PB-238 088

SUMMARY OF THE OFFICE OF TOXIC SUBSTANCES REQUIREMENTS RESULTING FROM THE TOXIC SUBSTANCES CONTROL ACT AND A PRELIMINARY SPECIFIC FOR A DATA MANAGEMENT SYSTEM

National Bureau of Standards

Prepared for
Environmental Protection Agency

August 1974

DISTRIBUTED BY:
NTIS
National Technical Information Service
U. S. DEPARTMENT OF COMMERCE

NBS-114A (REV. 7-73)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET	1. PUBLICATION OR REPORT NO.	2. Gov't Accession No. -	PB 238 088

4. TITLE AND SUBTITLE
Summary of the Office of Toxic Substances Requirements Resulting from the Toxic Substances Control Act and a Preliminary Specific for a Data Management System

5. Publication Date -

6. Performing Organization Code -

7. AUTHOR(S) John L. Berg, Josephine Walkowicz, Dennis Branstad, Michael Keplinger

8. Performing Organ. Report No.

9. PERFORMING ORGANIZATION NAME AND ADDRESS

NATIONAL BUREAU OF STANDARDS
DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20234

10. Project/Task/Work Unit No. 640 1411

11. Contract/Grant No. EPA-IAG-D4-0404

12. Sponsoring Organization Name and Complete Address (Street, City, State, ZIP)

Environmental Protection Agency
Office of Toxic Substances
401 M Street, S.W.
Washington, D. C. 20460

13. Type of Report & Period Covered
Final

14. Sponsoring Agency Code

15. SUPPLEMENTARY NOTES

Reproduced by
NATIONAL TECHNICAL INFORMATION SERVICE
US Department of Commerce
Springfield, VA. 22151

16. ABSTRACT *(A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)*

This report presents a requirements analysis and feasibility study for the data management system needed to use effectively industrial reporting data resulting from the proposed Toxic Substances Control Act.

The study finds that the Office of Toxic Substances requires a system with flexibility, extensibility of data content, ability to handle a wide and confidential nature of the reports, and suitability for immediate installation on a production basis.

In the study both a manual system that minimally satisfies the basic requirements and a computerized system with much extended capabilities are found technically feasible. In addition, the study presents feasible enhancements to the manual system which extend the manual system capabilities and show that a continuum of system decisions exists between the manual and the computerized system.

The study recommends immediate preparation for the computerized system in parallel with the adoption of a cost-saving manual system that has a four-year life expectancy. The manual system will provide the basis of the archival storage under the computerized system. Preparation for the eventual computerized system includes review of real experience under the manual system. It is not unlikely that real experience will lead to some revision in the projected six-year program.

17. KEY WORDS *(six to twelve entries; alphabetical order; capitalize only the first letter of the first key word unless a proper name; separated by semicolons)* Data Management System; Environmental Protection Agency; feasibility study; industrial reporting; Office of Toxic Substances; requirements analysis; toxic substances.

PRICES SUBJECT TO CHANGE

18. AVAILABILITY ☐ Unlimited

☐ For Official Distribution. Do Not Release to NTIS

☐ Order From Sup. of Doc., U.S. Government Printing Office Washington, D.C. 20402, SD Cat. No. C13

☒ Order From National Technical Information Service (NTIS) Springfield, Virginia 22151

19. SECURITY CLASS (THIS REPORT) UNCLASSIFIED	21. NO. OF PAGES 143
20. SECURITY CLASS (THIS PAGE) UNCLASSIFIED	22. Price 5.75

347025

PB-238 088

EPA-560/3-74-001

REPORT TO THE DIRECTOR,
OFFICE OF TOXIC SUBSTANCES

Summary of the Office of Toxic Substances Requirements resulting from the Toxic Substances Control Act and a preliminary specification for a data management system.

Prepared by

Institute for Computer Sciences and Technology

August, 1974

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
US Department of Commerce
Springfield, VA. 22151

ia

This report has been reviewed by the Office of Toxic Substances, EPA, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

ABSTRACT

This report presents a requirements analysis and feasibility study for the data management system needed to use effectively industrial reporting data resulting from the proposed Toxic Substances Control Act.

The study finds that the Office of Toxic Substances requires a system with flexibility, extensibility of data content, ability to handle a wide and confidential nature of the reports, and suitability for immediate installation on a production basis.

In the study both a manual system that minimally satisfies the basic requirements and a computerized system with much extended capabilities are found technically feasible. In addition, the study presents feasible enhancements to the manual system which extend the manual system capabilities and show that a continuum of system decisions exists between the manual and the computerized system.

The evaluation stresses that a cost-effective analysis of the two systems must include a comparison of the aids and services offered by the more expensive computerized system against the limited tools provided by the less expensive manual system. Further, the estimates provided for the six year period argue strongly that the annual report handling load and the accumulated archival data base will continue to grow beyond 1980. This growth will tax the limited capabilities of the manual system by

the fourth year but certain of the enhancements suggested can extend the utility of the manual system.

The study recommends immediate preparation for the computerized system in parallel with the adoption of a cost-saving manual system that has a four-year life expectancy. The manual system will provide the basis of the archival storage under the computerized system. Preparation for the eventual computerized system includes review of real experience under the manual system. It is not unlikely that real experience will lead to some revision in the projected six-year program.

EXECUTIVE SUMMARY OF MAJOR ISSUES

Manual or Automated Data Base

Discussion: Over the six year period studied, the annual report input estimates ranged from 10,500 per year (42 per day) in 1975 to an input of 90,500 (362 per day) in 1980. A simple comparison of costs between a manual system and automated system ignores the additional benefits achievable through automation. The manual system could only become increasingly cumbersome with the increasing reporting volume and file size. Further, a manual system would have limited response time and flexibility compared to the automated system. However, current planning necessarily reflects estimates. In addition to cost savings, the manual system would permit the gathering of real data and the development of better estimates using a minimal cost system. This data would assist not only the decision to automate but assist the many operational decisions OTS must make after deciding to automate.

Recommendation: OTS should initiate a manual system with an expected life span of four years and immediately begin preparation for a computerized report management system for operation in the fourth year. This would permit a parallel operation of both systems in the fourth year. At the same time, OTS should develop a formal program for monitoring actual experience with the reports and the OTS line units utilization of them. All planning should anticipate and prepare for eventual computerization of the OTS data base.

Confidentiality

Discussion: Considerations of confidentiality are pervasive in both the manual and automated system. In all cases and under any mode of ultimate implementation, confidentiality remains the responsibility of the Office of Toxic Substances. Confidentiality considerations include procedures related to data collection, entry, storage, processing, output, and dissemination. Factors affecting confidentiality considerations are personnel, media, facilities, equipment, administration, and quality control.

Recommendation: The Office of Toxic Substances should recognize the complexity of implementing confidentiality considerations within its data base and initiate an overall management oriented approach. This approach should develop a security program focused on specific problems and high risk areas. The Office of Toxic Substances should include confidentiality considerations in all phases of planning for the data base.

TABLE OF CONTENTS

Page

I. INTRODUCTION..1

II. REQUIREMENTS...2
  A. Office of Toxic Substances Requirements.......................2
    1. Reports...2
    2. Summary of Requirements.....................................4
    3. Interfaces with Existing data bases.........................5
    4. Chemical Substance Searching................................7
  B. Confidentiality...10
    1. Applicable Law and Legislative Requirements.................10
    2. Threats to Data...12
    3. Data Security and Confidentiality Considerations............13
    4. Summary...25
  C. System Considerations...27
    1. Reporting Form Structure....................................27
    2. Data and Report Volumes.....................................42
    3. OTS Use of the Data Base....................................49
III. FEASIBILITY OF ALTERNATIVE APPROACHES.........................51
  A. Manual System...51
    1. Information Flow..51
    2. Retrieval of Information....................................56
    3. Personnel Considerations....................................57
  B. Enhancements to the Manual System.............................63
    1. Description...63
    2. Indexing Aids...65
    3. Catalog posting and publication.............................65
    4. Data descriptor vacabulary..................................66
    5. Data Element Dictionary.....................................66
    6. Bibliography preparation....................................67
    7. Automated mailing list and tickler file.....................67
    8. Microfilm working files.....................................68
    9. Microfilm Retrieval Systems.................................68
    10. Personnel Considerations...................................68
  C. The Computerized System.......................................69
    1. Information Flow..71
    2. Data Entry Considerations...................................75
    3. Data Management System Selection............................79
    4. Data Management System Review...................... 80
    5. System Configuration.. 90
    6. Personnel Considerations....................................90
IV. EVALUATION AND COSTING CONSIDERATIONS..........................94
  A. Evaluation Considerations.....................................94
    1. Evaluation Criteria...94
    2. Facilities Evaluation.......................................95

- B. Comparative Cost/Benefit Analysis........................ 102
  - 1. Cost Estimates..102
  - 2. Comparative analysis..................................105
  - 3. Comparative Costs.....................................112
- C. Guidance for RFP..120

V. RECOMMENDATIONS...126
- A. Overview..126
- B. Recommended Program...127
  - 1. First year (1975).....................................127
  - 2. Second year (1976)....................................127
  - 3. Third year (1977).....................................129
  - 4. Fourth year (1978)....................................129
  - 5. Fifth and Sixth years (1979-1980).....................129
- C. Specific Topics...130
  - 1. Chemical substance searching.......................... 130
  - 2. Confidentiality.......................................130
  - 3. Reporting forms.......................................130
  - 4. Interface with existing data bases....................131

Appendix A. Existing Bibliographic Data Bases......................132

I. INTRODUCTION

As required under the Inter-agency Agreement between the National Bureau of Standards and the Environmental Protection Agency (EPA-IAG-D4-0404), this report presents a summary of the Office of Toxic Substances requirements and a specification for a data management system suitable for the technical basis of a Request for Proposals for systems design and implementation.

The report addresses those OTS information needs that satisfy the requirements of the proposed Toxic Substances Control Act. The report reflects interviews with several operating units within the Office of Toxic Substances and other government agencies. The report contains preliminary system considerations which identify the major components and feasible alternatives with associated cost estimates, a statement of the Office of Toxic Substances confidentiality requirements, and preliminary feasibility considerations to provide the technical basis of a Request for Proposals.

II. Requirements

A. Office of Toxic Substances Objectives and Information Requirements

The proposed data bank of manufacturer reports will need to be organized in a manner such that the reporting requirements imposed by legislation now pending in Congress can be satisfied in accordance with the provisions of the proposed Toxic Substances Act.

The new legislation will provide the basis for imposing various reporting and other obligations on manufacturers, importers, and processors of toxic substances as well as on the Administrator of EPA. Since the final form and date of enactment of the Toxic Substances Act are still unknown, requirements of the Act are summarized from the House version of the Act, Committee Print dated July 25, 1973.

1. Reports

Reporting requirements of manufacturers, processors, and importers are outlined below. "Manufacturer" as used subsequently in this report is used as a general term to include processors and importers as well, except in those cases where reporting requirements may differ among the three categories. The House Bill Section column cites the section containing the requirements.

a. Manufacturer Reports

(1) Annual Reports	House Bill Section
All Chemicals regulated by OTS	8(a)

Frequency: annually and at such more frequent times as the Administrator may require

Data: name, identity, categories of use amounts produced, and by-products of all chemical substances manufactured, as indicated in Exhibit 2.

(2) Premarket Reports

House Bill Section

(a) New Chemical Substances

Reporting time: 90 days in advance of commercial production of new chemical substance 5(d)

Data: same as a(1) above, and

: test data developed for intended use or distribution; or

(b) New Uses of Existing Substances

Reporting time: 90 days in advance of manufacture or distribution in commerce of significant new use of substance(s) listed in Federal Register or of existing chemical substance(s) 5(c)

Data: same as a(1) above

: test data developed for intended new use

(3) Other Manufacturer Reports

House Bill Section

(a) Performance of Tests

Frequency: as required or requested, for chemical substances to which a test protocol applies; 4(d)

: or petition Administrator for test protocol; 5(e)

: or share costs for performance of tests 5(f)

Type: test data developed under given protocol; or dollar share of costs

2. Summary of Requirements

a. Overall

- *The information handling capabilities of the Office of Toxic Substances data base system should offer sufficient flexibility to allow the inevitable evolution of requirements as more experience is gained.*
- *The Office of Toxic Substances storage and retrieval response time may be 48 to 72 hours.*
- *Reports submitted to the Office of Toxic Substances should be in the form of "finished" reports rather than simple computer printouts.*
- *Legal requirements necessitate the archival storage within EPA of signed reports as legal documents.*
- *In designing the information system, care should be taken to use only techniques well within the current state of the art.*
- *The original documents and all subsequent copies of the information contained therein (in whatever form that may be necessary) remain the exclusive property of the EPA.*
- *The system design should insure independence from any particular vendor.*
- *Correspondence to and from respondents should be handled by OTS personnel.*
- *Data declared by the respondent as confidential must be so marked on all reports.*

- *The information handling capabilities should be extensible in order to meet future as well as current needs.*

b. Usage Requirements

- *The manufacturer data must be readily accessible (with a minimum amount of clerical help) to users within the Office of Toxic Substances.*
- *When necessary, the user should be able to carry hard copy back to his desk for more detailed review without destroying the integrity of the data base for other users.*
- *The original documents must be preserved to satisfy legal requirements.*
- *The data base should support the identification of "Trend" information and of anomalies in reported production rates.*

3. Interfaces with Existing data bases.

As discussed elsewhere, the Office of Toxic Substances Data Base represents only a small part of the information needs of the Office. The expertise represented by the staff of OTS includes the knowledge of where to find the answers to questions raised by operational requirements of the Toxic Substances Control Act. The OTS data base will supplement the current sources of information and may even become a prime source that points to other data sources.

These other data sources are existing files of information in either manual or automated form. Over 175 such data files are reported in the Environmental Information Systems Directory published by the EPA.

The first step in accessing any file of information is through descriptions of the material available. The Systems Directory mentioned in the previous paragraph is an example of such a description. At a lower level, the description must provide an explanation of the information provided to describe each of the elements of the data files. At a still lower level, the description may require an explanation of the terms used in the various data elements, such as, the units used for numeric quantities.

In dealing with automated data files or data bases, the technical question of the interchangeability of data between the data bases is raised frequently. At a trivial level, data from existing automated data bases can be reduced to hard copy for inclusion within a manual data base. Alternatively, the existence of an automated data base can be indicated at the point in a manual system where any searcher for the pertinent information could find it.

The interchangeability of data between data files within the same retrieval system is often confused with the interchange of data between two computer systems. EPA with its Univac and IBM computers and S2000 Data Management System can support many automated data files, each with a content independent of all the others. Any user permitted by the system to access the data files can do so within the limits of the system itself. One such limit may be the lack of a capability to allow two files to be opened concurrently, so that data from one file can be compared to the other. Many systems lack this capability.

On the other hand, the interchange of data between dissimilar computer systems raises the question of incompatibility of the two systems' storage media, the codes used for storing information, the capacity of storage media, and similar incompatibilities resulting from physical differences between the two computer systems.

OTS's solution to the latter problem is to initiate development of data standards prior to acquiring physical or hardware system components. In the event that OTS seeks satisfaction of its needs through an outside service organization, data standards will still be necessary.

4. Chemical Substance Searching

OTS needs for chemical structure searching, under the proposed Toxic Substances Control Act, fall in the area of analysis of information reported by manufacturers rather than in the specification of the reporting and confidentiality requirements resulting from the proposed legislation. True, they are related. Chemical structure searching will not be possible unless chemical structure information is available. Summarized below are some points relevant to existing chemical structure techniques and OTS requirements in this area.

a. The Wiswesser notation and other linear notations are essentially "pre-computer" languages, not suitable for mechanized systems. [4]

b. The CAS Registry System, and systems based on CAS Registry Numbers, offer the best means for accessing existing stores of chemical information.

c. The technology for graphic displays and interactive searching capability on chemical structures is promising, but operational systems on large files are as yet not generally available.

d. The final legislation may not permit requiring structure information.

e. There is precedent for CAS cooperation both for registering previously unreported compounds and for maintaining confidentiality of information.

f. It is felt that chemical structure searching which is an analysis function within OTS cannot hope to be satisfied by the management information system established for accepting manufacturer reports. At best the "hooks" for linking with structure-oriented systems should be established within the management information system.

References

1. "Chemical Structure Handling, A Review of the Literature 1962-1968," National Academy of Sciences, Washington, D. C. 1969.
2. "Survey of Chemical Notation Systems," National Academy of Sciences, Washington, D. C. 1964.
3. "Progress Toward a Computer-Based Chemical Information System," Fred Tate, Chemical and Engineering News, Vol. 45, Jan. 1967.
4. "Interactive Graphic Chemical Structure Searching," Division of Computer Research and Technology, National Institutes of Health, May 1973.

B. Confidentiality

The Toxic Substances Control Act of 1973, S.426, among its provisions specifies the information that is to be collected, the reports which are to be generated, and confidentiality requirements for the data. Policy and other public law dictate further confidentiality considerations. The means of meeting confidentiality requirements in the design, implementation and operation of the OTS information system are presented below. The requirements are not unlike those found in other Federal agencies which deal with commercial trade secrets and similar data.

Providing the security necessary to protect confidential information is a complex task requiring an overall management-oriented approach. The security program must focus on specific problem areas of high risk but must yet remain a balanced overall program insuring certain levels of protection at all points in the information management process. Attention must be given to security at all times in the planning and development of the information system. Security plays a vital role regardless of whether the system is implemented as a manual, paritally conputerized, or fully automated.

1. Applicable Law and Legislative Requirements

Because the OTS system must deal with confidential and commercially sensitive information, strict adherence to statutory requirements and policy is absolutely essential. Certain legislation, particularly the Freedom of Information Act (5 U.S.C. 552), imposes requirements relating to both confidential and public information.

The Freedom of Information act makes provisions for public access to data maintained by Federal Agencies. That Act, however, exampts certain classes of information from mandatory disclosure requirements including information which contains or relates to a trade secret. Further, sanctions against Federal employees may be imposed under provisions of 18 U.S.C. 1905 (1948). These sanctions include a fine of not more than $1,000, or imprisonment for not more than one year or both, and removal from office or employment.*

The House version of S.426 contains provisions relating to confidentiality of data in §15. These provisions require that information which contains or relates to a trade secret or other matter referred to in §1905 or title 18, United States Code, shall be considered confidential and disclosed only under certain circumstances. The House version of S.426 specifies that such information may be made available to other officers or employees concerned with carrying out the Act (including the Chemical Substances Board and Committees formed under §10), or when relevant in any proceeding under the Act except that disclosure in such proceedings shall preserve the confidentiality to the extent possible without impairing the proceeding. Such information shall also be made available upon requests of duly authorized Committees of Congress. The confidentiality provisions of S.426 clearly indicate that OTS must ensure that the information processing system provide simple and efficient access to the information by authorized personnel yet control access and use of this information in a manner

---

**A trade secret may be defined as "any formula, pattern, device or compilation of information which is used in one's business, and which gives him an opportunity to obtain an advantage over competitors who do not know or use it. It may be a formula for a chemical compound..."[5].*

unauthorized by the Act.

2. Threats to Data

OTS must assume the responsibility for maintaining the confidentiality of data from collection through final analysis and dissemination, regardless of whether the data is processed manually or by a computer, within the Agency or under contract. Procedures must be implemented which insure that trade secret data is specifically identified and tagged at all stages in processing, especially during output, in order that no unauthorized dissemination is made of this data.

A risk analysis will provide the basis upon which a valid security program can be established. The recommended steps in this process include:

- An estimate of all potential damages resulting from (1) loss or destruction of data and program files; (2) theft of confidential or trade secret information; and (3) physical destruction or theft of physical information resources.
- An estimate of the probability of occurrence for potential threats and their effect on the information system in terms of the classes of loss potential.
- Development of an annual loss expectancy by combining the estimates of loss potential and threat probability.
- Selection of the array of remedial measures which effect the greatest reduction in the annual loss expectancy at the least total cost.

Detailed procedures for performing a risk analysis are contained in the Federal Information Processing Standard publication entitled, "Guidelines for ADP Physical Security and Risk Management"[11].

Application of risk management techniques to the design and development of the OTS information system will result in a comprehensive security plan *based on foreseeable threats against the security of the system. Once the threats are identified and an appropriate set of safeguards is selected to counter those threats, the implementation of safeguards can be assured whether the information system is developed in-house or by an outside contractor.*

Regardless of where the information system selected is implemented, the responsibility for assuring data confidentiality remains the primary function of OTS. This function must consequently be assigned a place in the administrative structure charged with the development of the system. Administrative functions and other data security considerations are addressed in the succeeding sections.

3. Data Security and Confidentiality: Considerations and Recommendations for Implementation

Figure 1 depicts the areas of primary consideration in planning for data security in information systems. This section discusses the relationships of the factors which have been checked in that matrix.

a. Administration

Specific security measures such as indoctrination materials and a security audit program can be developed either by OTS or by a contractor with OTS supervision. Security audit consists of producing a detailed record of the computer system's activity and analyzing this record, both manually and automatically, in order to signal possible security violations, produce security status summaries and generate damage assessment reports.

Control and application programs should maintain duplicate sets of data which should be separately stored and protected by management. Finally, automated techniques should implement the data security requirements for personal accountability for system usage and data access.

Administrative control of data and the people handling the data are the simplest and most economical methods of protecting data. These controls must be adequately specified and uniformly applied. Such controls include:

- *A precise designation of personnel who have access to the facilities, equipment, programs, storage media, and data;*
- *Precise specification of data access privileges: read, copy, modify and delete;*
- *Designation of a senior OTS staff member, whose duties include the role of security manager, independent of ADP operations, with direct responsibility and authority in security matters;*
- *Periodic independent internal security audits and aperiodic external audits of security procedures.*

Recommended administrative steps to organize the OTS security program include the following:

- *Assignment of OTS managerial responsibility for information security and establishment of a task force to prepare a security program plan.*
- *Performance of a preliminary risk analysis to identify major problem areas and select top priority security measures as needed to correct major problem areas.*
- *Initiation of a security program including:*

  Preparation of a plan and a schedule for implementing selected remedial measures identified in the preliminary risk analysis.

  Preparation and maintenance of a policy and plans handbook to include: (1) a security and data confidentiality policy statement; (2) mandatory security *procedures; and (3) guidelines for incorporating security measures in system design, to include programming, testing, and maintenance for the automated portions of the system.*

Internal procedural security should be implemented by:

- *Determination of potential targets for fraud, theft, or misuse of information resources by analyzing the work flow and the nature of the tasks performed by in-house or contractor personnel. This requires the development of procedures that will minimize exposure to loss of information resources. Such procedures may include (1) requiring cooperation between two individuals to perform critical tasks; (2) performance of additional checks and bounds comparisons; (3) formalization of standards for high risk operations; and (4) independent quality control checks.*
- *Designation of critical positions in the automated portion of the system for management, system programming, program library control, input/output control, exception processing, applications programming, data base management, quality control, internal audit, hardware maintenance, and definition of appropriate pre-employment screening requirements.*
- *Maintenance of up-to-date and accurate organization charts, delineations of responsibilities for functions performed either in-house or by contractor personnel, and work statements for appropriate key positions.*
- *To prevent unauthorized processing, implementation control of and record keeping procedures for job initiation, scheduling and distribution of output.*
- *Control of access to physical data files in order to maintain data integrity, to protect storage media, to provide audit trails of custody and use of data file, and to prevent unauthorized use of data files.*
- *Utilization of audit trails.*
- *Establishment of policy and procedures for program and data file retention to satisfy requirements for (1) backup operation; (2) audit and management review of operations; (3) control of program maintenance; (4) quality controls on input data; and (5) non-dependence of one individual's knowledge of systems and programs.*
- *Establishment of procedures to assure personal accountability for confidential information both in OTS and in contractor facilities.*

- *Assignment of several people to each sensitive task to reduce threats of collusion.*
- *Rotation of personnel among sensitive tasks.*

b. Personnel

Figure 1 shows that personnel security factors are pervasive throughout all information processing activities. These factors are especially important where there is close personal access to the information in human readable form. The most serious personnel threats to information are modification or disclosure, either accidentally or maliciously. In computer based information systems data is also vulnerable to personnel threats during processing and, to a lesser extent, while in storage.

OTS should develop selection criteria for all personnel who are involved in all stages of information processing identified Figure 1. The integrity of an employee is likely to be influenced by factors which can be evaluated before employment, including personal background, financial situation and motivation. Through initial orientation and periodic security indoctrination, personnel training and education should emphasize the importance of security procedures, in addition to adequately preparing the individual to perform assigned tasks correctly and efficiently.

Personal accountability and responsibility for confidential information is a key to ensuring data security. Automated techniques for satisfying this requirement when processing computerized data are subsequently discussed. Every employee handling confidential information must be aware that he is personally accountable for such information under his control and that he is responsible for its protection.

	Administration	Personnel	Media	Facilities	Equipment	Quality Control
Data Collection	✓	✓	✓			✓
Data Entry	✓	✓	✓	✓	✓	✓
Data Storage	✓	✓	✓	✓	✓	✓
Data Processing	✓	✓		✓	✓	✓
Data Output	✓	✓	✓	✓	✓	✓
Data Dissemination	✓	✓	✓			✓

Figure 1 Data Security Considerations

To minimize or obviate personnel factors which could contribute to compromise of data, the following steps are recommended:

- Determination of the security training requirements for OTS and contractor senior management, operations staff, support staff, etc.
- Selection and implementation of appropriate security awareness techniques such as training lectures and seminars; orientation booklets; amendments to job descriptions making employees responsible for security; and specification of individual responsibilities for data security.
- Development of personnel selection criteria that include consideration of factors such as personal background, financial situation and motivation.
- Dissemination of some (but obviously not all) of the security measures in force at the contractor facility as well as at OTS.
- Publicizing of selected cases of disclosure of confidential data at other installations when the penalties imposed were severe. Details of perpetration, however, should be omitted.
- Continuation of security awareness and education programs throughout system operation.

c. Media

The integrity of data to be processed under the provisions of S.426 is dependent on the types of media used for transmission and storage.

The media of data collection include physical media (paper form), the format of the data (representation, coding, description), and the method of transmission (postal service, courier). Confidentiality of the transmitted information must be ensured for all media employed. Vulnerability of unprocessed data to unauthorized access is perhaps highest during the collection phase. Data entry media involve an intermediate process (card punching, microfilming, key-to-tape, on-line keying) which transfers the data from its collection *media to computer entry or storage media. Data is particularly vulnerable to accidental modification during the entry phase. Data integrity assurance is discussed under quality control.*

Data storage media (punched cards, magnetic tape and disk, microforms, mass stores) have vastly different characteristics with respect to data security. Both archival storage and rapid access requirements must be considered. In particular, media degradation due to aging and heavy usage must be compared against anticipated data life and usage needs. (Geller discusses these considerations in reference [4].)

Data media for output and dissemination must be designed, procured, and protected. Paper forms used for report printing may be designed to facilitate analysis of the processed data and may be pre-printed with distribution and confidentiality instructions.

Recommendations under this heading include the following:

- *Adequate provisions for protection of source documents, input/output data, and programs throughout the entire processing cycle in the information facility, whether internal or contracted.*
- *Preprinting of forms with confidentiality instructions.*
- *Adequate provisions for protection of source documents data, and programs while in transit.*
- *Adoption of procedures to ensure that all trade secret data is specifically identified and tagged at all stages of processing.*
- *Maintenance of an inventory of the accession numbers and titles of report and data files in a location separate from the data itself.*

- Provisions of a capability for erasing magnetic media, either selectively or totally.

d. Facilities

Considerations for data processing facility security are important whether the processing is done by OTS or a contractor. Data processing facilities and equipment to be used are likely to vary greatly in specific proposals received from prospective vendors. Any contractual endeavor should emphasize the need for data security and specify the types of protection which will be provided by the contractor to satisfy OTS security requirements. Inadequate specification of security requirements may result in insufficient security to meet confidentiality needs, and, conversely, over specification may result in fewer responsive bids with higher costs.

Facilities are subject to environmental conditions (hurricane, tornado, flood, humidity, etc.), physical destruction (by fire, smoke, water, explosion, personal assault), and unauthorized personnel entry. Environmental and physical hazards may be anticipated and designed for, but they cannot be prevented. Contingency plans may be prepared, however, to minimize the effects of environmental losses. Unauthorized personnel entry can generally be prevented. Physical security of facilities should be specified for the processing of confidential, and in most cases all, data. Good engineering management should be evident in facility location, design, construction, and installation and should be considered in selecting contractors to process the OTS data. Personnel safety and comfort, utility availability, and physical access control should all be adequately engineered. Site preparation will vary greatly among candidate facilities for data entry, storage, processing and output; however, all must follow applicable guidelines for *safety and security. [11]*

Plans for security of the facility should include:

- *Selection of a facility or modification to an existing facility so that it can be easily protected against unauthorized entry.*
- *Minimizing threats from fire, water, wind, etc.*

Provisions for physical protection should include the following:

- *Identification of critical ADP areas to include the computer room, the mechanical equipment room, data control and conversion area, data file storage area, programmer's area, forms storage area, and provisions of adequate physical protection and access control for each area.*
- *Protection against theft, vandalism, sabotage, and forced intrusions by use of adequate lighting, intrusion detection systems, physical barriers at doors, windows, and other openings, and guards as required.*
- *Control of access to critical areas and ADP facilities by installation of conventional or electronic door locks, supervision by guards or receptionists over movement of people and materials, administrative procedures (sign-in logs, identification cards or badges, property passes and shipping receiving forms), and similar measures.*

e. Equipment

The equipment employed in data processing will have a significant effect on the ease of providing for data security. Since data security includes preventing accidental modification, the adequacy of equipment and its state of maintenance should be considered. Equipment managers may perform maintenance locally, contract for preventative maintenance and repair from the original vendor, or simply request ad-hoc repair from the supplier when the equipment fails. Data entry equipment (card punches and readers, teleprinters, CRT terminals, optical scanners, microfilm readers) *are combinations of mechanical and electronic machines. Even though technology has improved both disciplines, electronics has far surpassed mechanics in equipment reliability. Data integrity will depend on the design and condition of the data entry equipment and the man-machine interface between the operator and the equipment.*

Data storage equipment varies in the protection provided to stored data through the physical and electronic data handling techniques employed: reading and writing techniques, error detection/correction techniques, and storage media handling techniques. In addition, data storage equipment and facilities include those which store removable media. Tape libraries, card and forms storage, etc., must be physically protected from theft of the media and must be provided with fire, water, temperature, and humidity detection and control facilities. Provision should be made for cleaning, copying, and degaussing (magnetic erasure) equipment to be used for erasing storage media no longer required for archival magnetic storage.

Data security is of lesser concern in output equipment, unless this equipment is also the dissemination system (remote terminals, computer and terminal networks, telecommunications). Impact printers that are local to the data processing facility may be physically protected against unauthorized viewing of the printing. Data dissemination through reports is generally a manual task and hence regulated by administrative controls and procedures.

Data processing equipment and control and application programs which process the data should have a minimal set of safeguards incorporated in them, including means of user identification and verification, and isolation of specific data from users not authorized to access it. The implementation of these safeguards will vary among systems. Isolation techniques may range from dedicated use of the system by one user, to a very complex memory segregation algorithm. Personnel identification may be performed via personal recognition, badges, ID cards, or individual passwords.

Equipment security recommendations include:

- Selection of a computer system configuration which can operate at the required security level. Factors to consider in this section include the availability of operating systems incorporating state-of-the-art terminal access to the system and data files. Such a selection may require use of a dedicated computer system or use of a general system operated in a restricted manner (e.g., no user programming concurrent to sensitive processing).
- Utilization of user and data isolation capabilities in the system (memory protection and segmentation, data storage segregation).
- Implementation of a computerized authorization mechanism for controlling access to sensitive data files. Data authorization should include explicit data handling permissions such as read only, read/write, append, and delete.
- Adequate provisions for equipment maintenance and backup.

f. Quality Control

The implementation of quality control procedures throughout all information processing operations is a fundamental component of a data security program. Data integrity is a necessary prerequisite to the quality control process and is variously defined to mean correct data or data that is both correct and uncompromised. Adequate quality control procedures must be implemented to ensure that the data collected and entered is accurate, and to detect and prevent unauthorized or accidental modifications *to the data. Quality control procedures should include provisions to assure verification of the accuracy and integrity of the data at all steps in the information system. Assurance of integrity may include use of redundant storage and processing facilities. To catch errors introduced by faulty machines or noisy communication channels, extremely important data may be transmitted twice and the two transmission compared whether this transmission is within a single computer system or among a plurality of systems. Encryption techniques may be employed if transmission of data that is highly sensitive to disclosure is contemplated.*

Quality control procedures should be developed to provide for:

- *Establishment of an internal audit team to audit either OTS or contractor data processing procedures with representatives from the EPA audit, building safety and security, ADP unit, and users units.*
- *An audit plan and schedule for systematic validation of all critical security and emergency measures and reports to OTS project management.*
- *An audit reporting system to monitor quality control procedures in order to detect emerging deficiencies and to assure their prompt correction. A check list may be used for this purpose.*
- *Establishment of certification procedures to determine that security features of the system comply with OTS specifications which in turn satisfy the security requirements. Certification is required in order to check that the design is complete; to confirm that the implementation is correct; to determine that the installation meets all design standards and requirements; to establish that a system is secure after system modification, or failure, or after penetration (either actual or suspected) has taken place.*
- *Establishment of procedures to resecure the system if a penetration does take place, and to verify that the data base and processing programs are unaffected.*

4. Summary

This Section has presented the basic considerations of data confidentiality which OTS msut take into account under the provisions of the pending legislation. The trade-offs among generality, costs, efficiency, and security must be made when specifying initial conditions of contracts in data processing and in evaluating various responses from equipment and services suppliers. Defining and understanding the final goals of envrionmental protection will lead to realistic predictions of the needed results of data processing and hence to the data collection and processing requirements. This Section presented recommendations which may be implemented by OTS to effectuate these basic considerations.

References and Suggested Reading

(1) Branstad, Dennis K. and Susan K. Reed, "Executive Guide to Computer Security", U.S. Department of Commerce, National Bureau of Standards, Washington, D.C., NBS Special Publication, 1974.

(2) Davis, Ruth M., "Privacy and Security in Computer Systems: An Overview", CBEMA Privacy Series 2, Computer and Business Equipment Manufacturers Association, Washington, D.C., February 1974, 21 p.

(3) Geller, Sydney B., "The Effects of Magnetic Fields of Magnetic Storage Media Used in Computers", U.S. Department of Commerce, National Bureau of Standards, Washington, D.C., NBS Technical Note 735, July 1972, 30 p.

(4) Geller, Sydney B., "Factors in Archival Data Storage Systems", U.S. Department of Commerce, National Bureau of Standards, Washington, D.C., Working Paper submitted to Datamation, 1974, 16 p.

(5) Milgrim, Roger M., "Trade Secrets", Mathew Bender, New York, N.Y., 1972.

(6) Parker, Donn B., Susan Nycum and S. Stephen Oura, "Computer Abuse", Stanford Research Institute, Menlo Park, California, 1973, 131 p.

(7) "Records, Computers and the Rights of Citizens, Report of the Secretary's Advisory Committee on Automated Personal Data Systems", U.S. Department of Health, Education and Welfare, Washington, D.C., July 1973, 346 p.

(8) Reed, Susan K. and Dennis K. Branstad, "Controlled Accessibility Workshop Report", U.S. Department of Commerce, National Bureau of Standards, Washington, D.C., NBS Technical Note 827, May 1974, 82 p.

(9) Reed, Susan K. and Martha M. Gray, "Controlled Accessibility Bibliography", U.S. Department of Commerce, National Bureau of Standards, Washington, D.C., NBS Technical Note 780, June 1973, 11 p.

(10) Renninger, Clark R. and Dennis K. Branstad, "Government Looks at Privacy and Security in Computer Systems", U.S. Department of Commerce, National Bureau of Standards, Washington, D.C., NBS Technical Note 809, February, 1974, 37 p.

(11) "Risk Management and Physical Security", U.S. Department of Commerce, National Bureau of Standards, Washington, D.C., FIPS PUB 31.

C. System Considerations

In order to meet the requirements posed by the new legislation, OTS will need to seek services encompassing: (1) design and implementation of a manufacturer reporting system, as required by S.426; (2) a self-contained information management system to assist the Office of Toxic Substances in the control and management of the data base that will be suitable for storage, retrieval and reporting of data generated by the manufacturer reporting system; and (3) establishment and operation of a facility for processing and safeguarding confidentiality of data reported.

A candidate system must be flexible to evolve gradually as required to accommodate an initial volume of 10,000 manufacturer reports annually on approximately 500 chemical substances. As the initial volume of reports grows and as OTS information needs become more crystallized, the Chemical Abstracts file may be needed for chemical structure searching.

Some indeterminate proportion of the manufacturer report file will be involved in administrative transactions that will trigger signals for organizational units within OTS/EPA to take appropriate action: prepare a report, publish a notice in the Federal Register, etc.

1. Reporting Form Structure

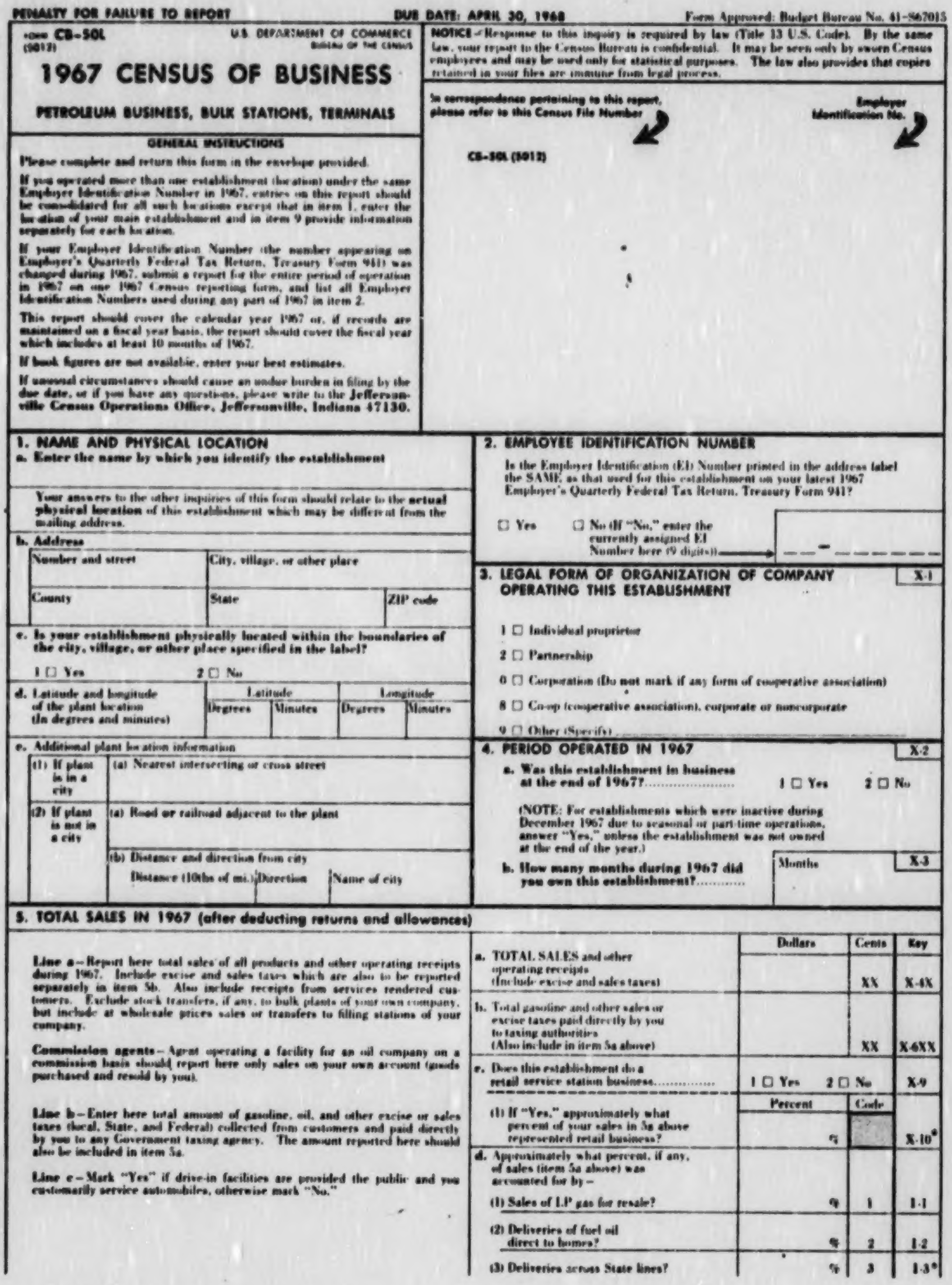
The actual design of the form to be used by the Office of Toxic Substances requires complete information about the eventual reporting regulations that will be issued by OTS. Therefore, the final design of the report form is left as a task for the providers of the data base services. In this discussion, some statements of OTS requirements are expressed and sufficient detail about the reporting form defined in order to provide a basis for the system description and cost estimates provided in this report.

The Bureau of Census form CB-50L (Exhibit 1-1) provides an excellent example of a reporting form to be considered by the Office of Toxic Substances. There is a great deal of information on the form that assists the respondent in understanding how, when, and to whom he should submit the forms. Similar information should be provided in the OTS form, so that respondents will have clear instructions on what data to report, what codes to use, etc. Exhibit 2 shows the data elements to be included in the reporting form, presented are five items with several fields in each item. The items are: manufacturer data; chemical substances; end-uses; by-products; and production features.

Manufacturer data is that portion of the reporting form that contains all the information needed to identify the manufacturer, processor, or importer who has submitted the report. There is only one such item needed for each annual report. Further, it can be assumed that subsequent years' reports will contain substantially identical information.

Chemical substance items contain information about a specific chemical substance. Like the information identifying the manufacturer, the information about the chemical substance can be assumed to remain constant over several annual reports.

EXHIBIT 1-1 SAMPLE CENSUS FORM



PENALTY FOR FAILURE TO REPORT **DUE DATE: APRIL 30, 1968** Form Approved: Budget Bureau No. 41-S67015

FORM **CB-50L**
(5012)

U.S. DEPARTMENT OF COMMERCE
BUREAU OF THE CENSUS

1967 CENSUS OF BUSINESS

PETROLEUM BUSINESS, BULK STATIONS, TERMINALS

NOTICE – Response to this inquiry is required by law (Title 13 U.S. Code). By the same law, your report to the Census Bureau is confidential. It may be seen only by sworn Census employees and may be used only for statistical purposes. The law also provides that copies retained in your files are immune from legal process.

In correspondence pertaining to this report, please refer to this Census File Number

Employer Identification No.

CB-50L (5012)

GENERAL INSTRUCTIONS

Please complete and return this form in the envelope provided.

If you operated more than one establishment (location) under the same Employer Identification Number in 1967, entries on this report should be consolidated for all such locations except that in item 1, enter the location of your main establishment and in item 9 provide information separately for each location.

If your Employer Identification Number (the number appearing on Employer's Quarterly Federal Tax Return, Treasury Form 941) was changed during 1967, submit a report for the entire period of operation in 1967 on one 1967 Census reporting form, and list all Employer Identification Numbers used during any part of 1967 in item 2.

This report should cover the calendar year 1967 or, if records are maintained on a fiscal year basis, the report should cover the fiscal year which includes at least 10 months of 1967.

If book figures are not available, enter your best estimates.

If unusual circumstances should cause an undue burden in filing by the **due date**, or if you have any questions, please write to the **Jeffersonville Census Operations Office, Jeffersonville, Indiana 47130.**

1. NAME AND PHYSICAL LOCATION

a. Enter the name by which you identify the establishment

Your answers to the other inquiries of this form should relate to the **actual physical location** of this establishment which may be different from the mailing address.

b. Address

Number and street	City, village, or other place	
County	State	ZIP code

c. Is your establishment physically located within the boundaries of the city, village, or other place specified in the label?

1 ☐ Yes 2 ☐ No

d. Latitude and longitude of the plant location (In degrees and minutes)	Latitude		Longitude	
	Degrees	Minutes	Degrees	Minutes

e. Additional plant location information

(1) If plant is in a city	(a) Nearest intersecting or cross street		
(2) If plant is not in a city	(a) Road or railroad adjacent to the plant		
	(b) Distance and direction from city Distance (10ths of mi.)	Direction	Name of city

2. EMPLOYEE IDENTIFICATION NUMBER

Is the Employer Identification (EI) Number printed in the address label the SAME as that used for this establishment on your latest 1967 Employer's Quarterly Federal Tax Return, Treasury Form 941?

☐ Yes ☐ No (If "No," enter the currently assigned EI Number here (9 digits)) → __ – _______

3. LEGAL FORM OF ORGANIZATION OF COMPANY OPERATING THIS ESTABLISHMENT

X-1

1 ☐ Individual proprietor

2 ☐ Partnership

0 ☐ Corporation (Do not mark if any form of cooperative association)

8 ☐ Co-op (cooperative association), corporate or noncorporate

9 ☐ Other (Specify) __________

4. PERIOD OPERATED IN 1967

X-2

a. Was this establishment in business at the end of 1967?.................. 1 ☐ Yes 2 ☐ No

(NOTE: For establishments which were inactive during December 1967 due to seasonal or part-time operations, answer "Yes," unless the establishment was not owned at the end of the year.)

b. How many months during 1967 did you own this establishment?............ Months X-3

5. TOTAL SALES IN 1967 (after deducting returns and allowances)

Line a – Report here total sales of all products and other operating receipts during 1967. Include excise and sales taxes which are also to be reported separately in item 5b. Also include receipts from services rendered customers. Exclude stock transfers, if any, to bulk plants of your own company, but include at wholesale prices sales or transfers to filling stations of your company.

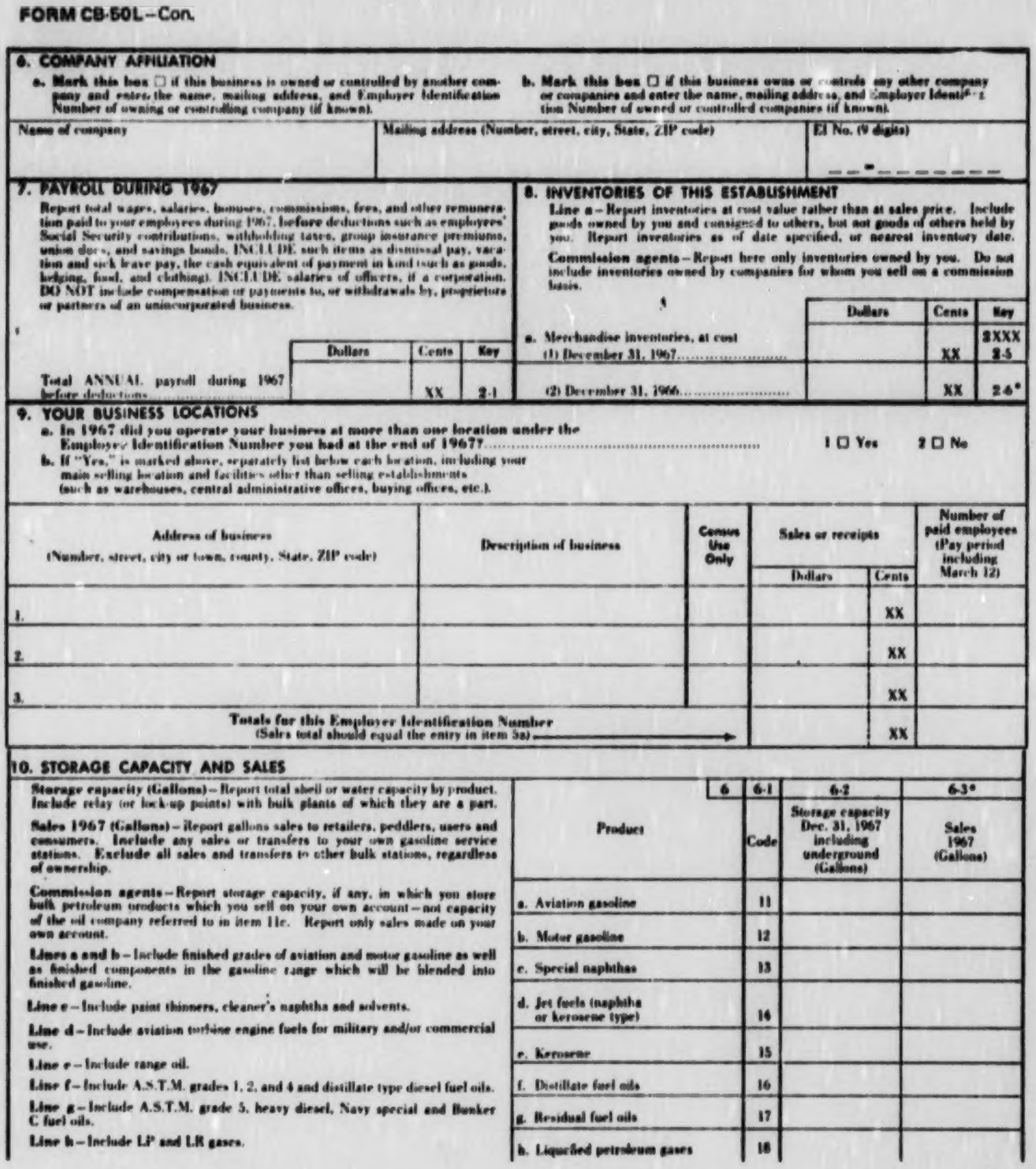
Commission agents – Agent operating a facility for an oil company on a commission basis should report here only sales on your own account (goods purchased and resold by you).

Line b – Enter here total amount of gasoline, oil, and other excise or sales taxes (local, State, and Federal) collected from customers and paid directly by you to any Government taxing agency. The amount reported here should also be included in item 5a.

Line c – Mark "Yes" if drive-in facilities are provided the public and you customarily service automobiles, otherwise mark "No."

	Dollars	Cents	Key
a. TOTAL SALES and other operating receipts (Include excise and sales taxes)		XX	X-4X
b. Total gasoline and other sales or excise taxes paid directly by you to taxing authorities (Also include in item 5a above)		XX	X-6XX
c. Does this establishment do a retail service station business................	1 ☐ Yes	2 ☐ No	X-9
	Percent	Code	
(1) If "Yes," approximately what percent of your sales in 5a above represented retail business?	%		X-10*
d. Approximately what percent, if any, of sales (item 5a above) was accounted for by –			
(1) Sales of LP gas for resale?	%	1	1-1
(2) Deliveries of fuel oil direct to homes?	%	2	1-2
(3) Deliveries across State lines?	%	3	1-3*

EXHIBIT 1-2 SAMPLE CENSUS FORM



FORM CB-50L–Con.

6. COMPANY AFFILIATION

a. **Mark this box** ☐ if this business is owned or controlled by another company and enter the name, mailing address, and Employer Identification Number of owning or controlling company (if known).

b. **Mark this box** ☐ if this business owns or controls any other company or companies and enter the name, mailing address, and Employer Identification Number of owned or controlled companies (if known).

Name of company	Mailing address (Number, street, city, State, ZIP code)	EI No. (9 digits)
		_ _ - _ _ _ _ _ _ _

7. PAYROLL DURING 1967

Report total wages, salaries, bonuses, commissions, fees, and other remuneration paid to your employees during 1967, **before** deductions such as employees' Social Security contributions, withholding taxes, group insurance premiums, union dues, and savings bonds. INCLUDE such items as dismissal pay, vacation and sick leave pay, the cash equivalent of payment in kind (such as goods, lodging, food, and clothing). INCLUDE salaries of officers, if a corporation. DO NOT include compensation or payments to, or withdrawals by, proprietors or partners of an unincorporated business.

	Dollars	Cents	Key
Total ANNUAL payroll during 1967 before deductions		XX	2-1

8. INVENTORIES OF THIS ESTABLISHMENT

Line a – Report inventories at cost value rather than at sales price. Include goods owned by you and consigned to others, but not goods of others held by you. Report inventories as of date specified, or nearest inventory date.

Commission agents – Report here only inventories owned by you. Do not include inventories owned by companies for whom you sell on a commission basis.

	Dollars	Cents	Key
a. Merchandise inventories, at cost (1) December 31, 1967		XX	2XXX 2-5
(2) December 31, 1966		XX	2-6*

9. YOUR BUSINESS LOCATIONS

a. **In 1967 did you operate your business at more than one location under the Employer Identification Number you had at the end of 1967?** 1 ☐ Yes 2 ☐ No

b. If "Yes," is marked above, separately list below each location, including your main selling location and facilities other than selling establishments (such as warehouses, central administrative offices, buying offices, etc.).

Address of business (Number, street, city or town, county, State, ZIP code)	Description of business	Census Use Only	Sales or receipts		Number of paid employees (Pay period including March 12)
			Dollars	Cents	
1.				XX	
2.				XX	
3.				XX	
Totals for this Employer Identification Number (Sales total should equal the entry in item 5a) →				XX	

10. STORAGE CAPACITY AND SALES

Storage capacity (Gallons) – Report total shell or water capacity by product. Include relay (or lock-up points) with bulk plants of which they are a part.

Sales 1967 (Gallons) – Report gallons sales to retailers, peddlers, users and consumers. Include any sales or transfers to your own gasoline service stations. Exclude all sales and transfers to other bulk stations, regardless of ownership.

Commission agents – Report storage capacity, if any, in which you store bulk petroleum products which you sell on your own account – not capacity of the oil company referred to in item 11c. Report only sales made on your own account.

Lines a and b – Include finished grades of aviation and motor gasoline as well as finished components in the gasoline range which will be blended into finished gasoline.

Line c – Include paint thinners, cleaner's naphtha and solvents.

Line d – Include aviation turbine engine fuels for military and/or commercial use.

Line e – Include range oil.

Line f – Include A.S.T.M. grades 1, 2, and 4 and distillate type diesel fuel oils.

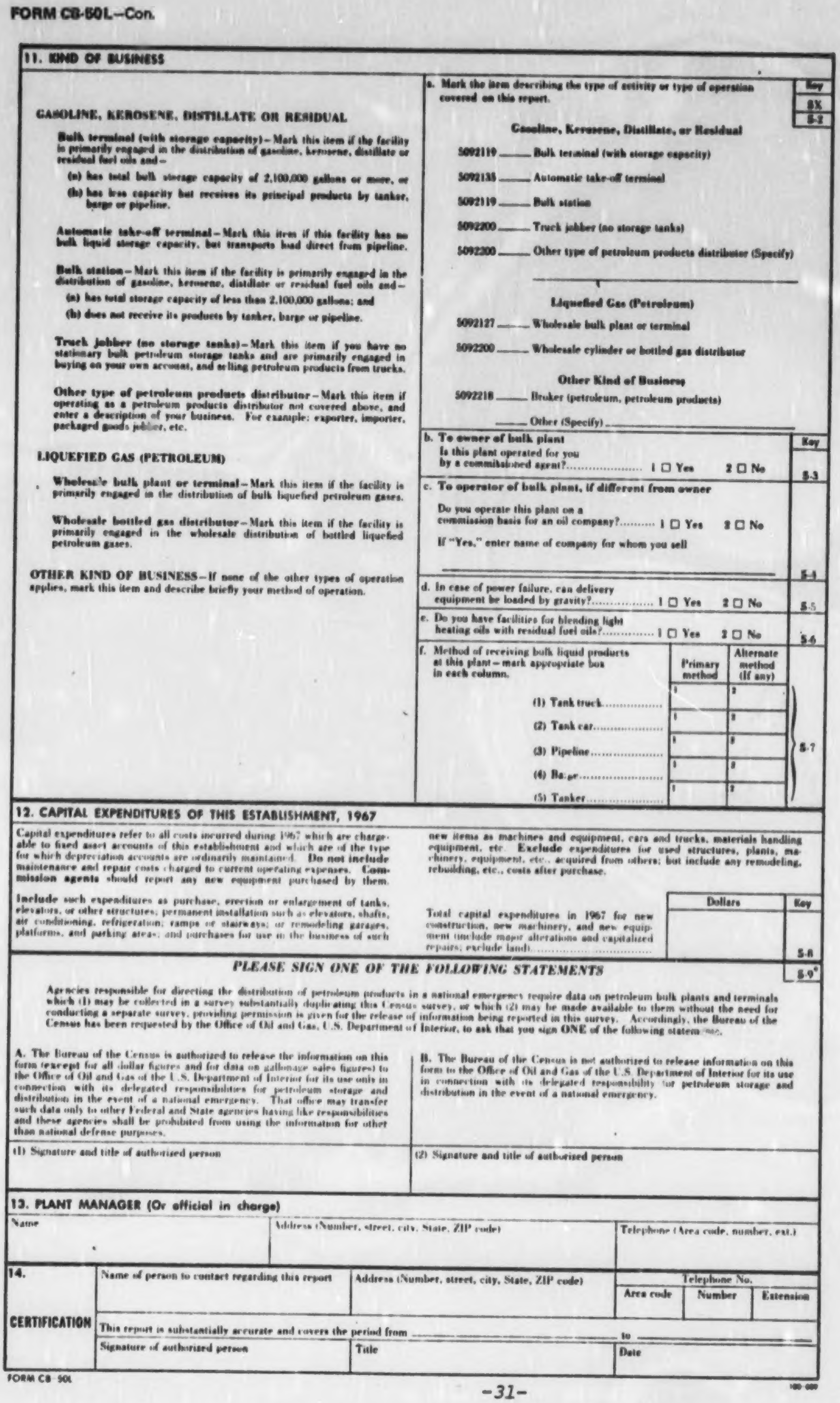
Line g – Include A.S.T.M. grade 5, heavy diesel, Navy special and Bunker C fuel oils.

Line h – Include LP and LR gases.

6	6-1	6-2	6-3*
Product	Code	Storage capacity Dec. 31, 1967 including underground (Gallons)	Sales 1967 (Gallons)
a. Aviation gasoline	11		
b. Motor gasoline	12		
c. Special naphthas	13		
d. Jet fuels (naphtha or kerosene type)	14		
e. Kerosene	15		
f. Distillate fuel oils	16		
g. Residual fuel oils	17		
h. Liquefied petroleum gases	18		

Reproduced from best available copy.

EXHIBIT 1-3 SAMPLE CENSUS FORM



FORM CB-50L—Con.

11. KIND OF BUSINESS

GASOLINE, KEROSENE, DISTILLATE OR RESIDUAL

Bulk terminal (with storage capacity)—Mark this item if the facility is primarily engaged in the distribution of gasoline, kerosene, distillate or residual fuel oils and—

(a) has total bulk storage capacity of 2,100,000 gallons or more, or

(b) has less capacity but receives its principal products by tanker, barge or pipeline.

Automatic take-off terminal—Mark this item if this facility has no bulk liquid storage capacity, but transports load direct from pipeline.

Bulk station—Mark this item if the facility is primarily engaged in the distribution of gasoline, kerosene, distillate or residual fuel oils and—

(a) has total storage capacity of less than 2,100,000 gallons; and

(b) does not receive its products by tanker, barge or pipeline.

Truck jobber (no storage tanks)—Mark this item if you have no stationary bulk petroleum storage tanks and are primarily engaged in buying on your own account, and selling petroleum products from trucks.

Other type of petroleum products distributor—Mark this item if operating as a petroleum products distributor not covered above, and enter a description of your business. For example: exporter, importer, packaged goods jobber, etc.

LIQUEFIED GAS (PETROLEUM)

Wholesale bulk plant or terminal—Mark this item if the facility is primarily engaged in the distribution of bulk liquefied petroleum gases.

Wholesale bottled gas distributor—Mark this item if the facility is primarily engaged in the wholesale distribution of bottled liquefied petroleum gases.

OTHER KIND OF BUSINESS—If none of the other types of operation applies, mark this item and describe briefly your method of operation.

a. Mark the item describing the type of activity or type of operation covered on this report.

Key: 8X S-2

Gasoline, Kerosene, Distillate, or Residual

5092119 ____ Bulk terminal (with storage capacity)
5092135 ____ Automatic take-off terminal
5092119 ____ Bulk station
5092200 ____ Truck jobber (no storage tanks)
5092200 ____ Other type of petroleum products distributor (Specify)
____

Liquefied Gas (Petroleum)

5092127 ____ Wholesale bulk plant or terminal
5092200 ____ Wholesale cylinder or bottled gas distributor

Other Kind of Business

5092218 ____ Broker (petroleum, petroleum products)
____ Other (Specify) ________________________________

		Key
b. To owner of bulk plant Is this plant operated for you by a commissioned agent?	1 ☐ Yes 2 ☐ No	S-3
c. To operator of bulk plant, if different from owner Do you operate this plant on a commission basis for an oil company?	1 ☐ Yes 2 ☐ No	
If "Yes," enter name of company for whom you sell ________________		S-4
d. In case of power failure, can delivery equipment be loaded by gravity?	1 ☐ Yes 2 ☐ No	S-5
e. Do you have facilities for blending light heating oils with residual fuel oils?	1 ☐ Yes 2 ☐ No	S-6

f. Method of receiving bulk liquid products at this plant—mark appropriate box in each column.	Primary method	Alternate method (If any)	
(1) Tank truck	1	2	S-7
(2) Tank car	1	2	
(3) Pipeline	1	2	
(4) Barge	1	2	
(5) Tanker	1	2	

12. CAPITAL EXPENDITURES OF THIS ESTABLISHMENT, 1967

Capital expenditures refer to all costs incurred during 1967 which are chargeable to fixed asset accounts of this establishment and which are of the type for which depreciation accounts are ordinarily maintained. **Do not include** maintenance and repair costs charged to current operating expenses. **Commission agents** should report any new equipment purchased by them.

Include such expenditures as purchase, erection or enlargement of tanks, elevators, or other structures; permanent installation such as elevators, shafts, air conditioning, refrigeration; ramps or stairways; or remodeling garages, platforms, and parking areas; and purchases for use in the business of such new items as machines and equipment, cars and trucks, materials handling equipment, etc. **Exclude** expenditures for used structures, plants, machinery, equipment, etc., acquired from others; but include any remodeling, rebuilding, etc., costs after purchase.

	Dollars	Key
Total capital expenditures in 1967 for new construction, new machinery, and new equipment (include major alterations and capitalized repairs; exclude land)		S-8

PLEASE SIGN ONE OF THE FOLLOWING STATEMENTS S-9*

Agencies responsible for directing the distribution of petroleum products in a national emergency require data on petroleum bulk plants and terminals which (1) may be collected in a survey substantially duplicating this Census survey, or which (2) may be made available to them without the need for conducting a separate survey, providing permission is given for the release of information being reported in this survey. Accordingly, the Bureau of the Census has been requested by the Office of Oil and Gas, U.S. Department of Interior, to ask that you sign **ONE** of the following statements.

A. The Bureau of the Census is authorized to release the information on this form (except for all dollar figures and for data on gallonage sales figures) to the Office of Oil and Gas of the U.S. Department of Interior for its use only in connection with its delegated responsibilities for petroleum storage and distribution in the event of a national emergency. That office may transfer such data only to other Federal and State agencies having like responsibilities and these agencies shall be prohibited from using the information for other than national defense purposes.

(1) Signature and title of authorized person

B. The Bureau of the Census is not authorized to release information on this form to the Office of Oil and Gas of the U.S. Department of Interior for its use in connection with its delegated responsibility for petroleum storage and distribution in the event of a national emergency.

(2) Signature and title of authorized person

13. PLANT MANAGER (Or official in charge)

Name	Address (Number, street, city, State, ZIP code)	Telephone (Area code, number, ext.)

14. CERTIFICATION

Name of person to contact regarding this report	Address (Number, street, city, State, ZIP code)	Telephone No. Area code	Number	Extension

This report is substantially accurate and covers the period from ______________ to ______________

Signature of authorized person	Title	Date

FORM CB-50L

Exhibit 2-1

Manufacturer Report
Manufacturer Item

Data Element Name	Field Type	Field Length
Corporate Name	AN	30*
Corporate Address:		
Street Number	AN	22*
Street Name		
City	A	20*
State	A	2*
Zip Code	N	5
County	A	15*
Contact Name	AN	20*
Contact Phone No.	AN	10
Reporting Estab. Name	AN	30*
Address:		
Street Number	AN	22*
Street Name		
City	A	20*
State	A	2
Zip Code	N	5
County	A	15*
Contact Name	AN	20*
Contact Phone No.	AN	10*
Corporate ID No.	N	9
Function:		
Manufacturer		
Processor	A	3
Importer		
Date Submitted	N	6 1/
Data Restriction Indicators	N	20 2/
Report No.	AN	10 1/
Test Data ID	AN	20 1/
TOTAL CHARACTERS		316

*Average Length

1/ To be supplied by OTS.

2/ It is assumed that an average of 20 boxes will be checked as confidential, trade secret, etc.

Exhibit 2-2
Manufacturer Report
Chemical Substance Item

Data Element Name	Field Type	Field Length
Substance Common Name	A	30*
CAS Registry Number	N	9
CAS Name	AN	30*
Trade Names:		
Trade Name 1	AN	30*
Trade Name 2	AN	30*
Trade Name 3	AN	30*
Synonym 1	AN	30*
Synonym 2	AN	30*
Synonym 3	AN	30*
Census Material Code	N	9
Two-D Representation	Graphic	
Molecular Weight	N	6
Test Protocol Available?	A	1 1/
TOTAL CHARACTERS		265

*Average Length

1/ To be supplied by OTS.

Exhibit 2-3
Manufacturer Report
Production Item

Data Element Name	*Field Type*	*Field Length*
Manufactured Quantities:		
1st Quarter	*N*	*9*
2nd Quarter	*N*	*9*
3rd Quarter	*N*	*9*
4th Quarter	*N*	*9*
Total Annual	*N*	*10*
Processed Quantities:		
1st Quarter	*N*	*9*
2nd Quarter	*N*	*9*
3rd Quarter	*N*	*9*
4th Quarter	*N*	*9*
Total Annual	*N*	*10*
Imported Quantities:		
1st Quarter	*N*	*9*
2nd Quarter	*N*	*9*
3rd Quarter	*N*	*9*
4th Quarter	*N*	*9*
Total Annual	*N*	*10*
TOTAL CHARACTERS		*138*

Comment: *It should be noted that not all of the above fields will contain data. An establishment that manufactures only, will obviously not report production data for processed or imported quantities.*

Exhibit 2-4
Manufacturer Report
End Uses Item

Data Element Name	*Field Type*	*Field Length*
SIC Code	*N*	*9*
Description	*A*	*100*
Annual Production	*N*	*10*
TOTAL CHARACTERS		*119*

X 25 end uses = 2,975 characters/chemical substance

Note: It is estimated that an average of 25 end uses will be reported for each chemical substance.

Exhibit 2-5
Manufacturer Report
By-products Item

Data Element Name	Field Type	Field Length
By-products:		
Chemical Name	A	30*
Annual Production	N	9
Disposal Method	A	30*
By-product of Use?	A	1
By-product of Disposal?	A	1
By-product of Manufacture?	A	1
Effluent - Gallons per Day	N	8
	TOTAL CHARACTERS	80

X 4 by-products = 320 characters/chemical substance

Note: It is estimated that 4 by-products will be reported for each chemical substance.

* Average Length

Exhibit 2-6
Premarket Report
Manufacturer Item

Data Element Name	Field Type	Field Length
Corporate Name	AN	30*
Corporate Address:		
Street Number Street Name	AN	22*
City	A	20*
State	A	2
Zip Code	N	5
County	A	15*
Contact Name	AN	20*
Contact Phone No.	AN	10
Reporting Estab. Name	AN	30*
Address:		
Street Number Street Name	AN	22*
City	A	20*
State	A	2
Zip Code	N	5
County	A	15*
Contact Name	AN	20*
Contact Phone No.	AN	10*
Corporate ID No.	N	9
Function: Manufacturer Processor Importer	A	3
Date Submitted	N	6 1/
Data Restriction Indicators	N	20 2/
Report No.	AN	10 1/
Test Data ID	AN	20 1/
TOTAL CHARACTERS		316

*Average Length

1/ To be supplied by OTS.

2/ It is assumed that an average of 20 boxes will be checked as confidential, trade secret, etc.

Exhibit 2-7
Premarket Report
Chemical Substance Item

Data Element Name	Field Type	Field Length
Substance Common Name	A	30*
CAS Registry Name	N	9
CAS Name	AN	30*
Trade Names:		
Trade Name 1	AN	30*
Trade Name 2	AN	30*
Trade Name 3	AN	30*
Synonym 1	AN	30*
Synonym 2	AN	30*
Synonym 3	AN	30*
Census Material Code	N	9
Two-D Representation	Graphic	
Molecular Weight	N	6
Test Protocol Available?	A	1 1/
TOTAL CHARACTERS		265

*Average Length

1/ To be supplied by OTS.

Exhibit 2-8
Premarket Report
By-products Item

Data Element Name	Field Type	Field Length
By-products:		
Chemical Name	A	30*
Annual Production	N	9
Disposal Method	A	30*
By-product of Use?	A	1
By-product of Disposal?	A	1
By-product of Manufacture?	A	1
Effluent - Gallons per Day	N	8
	TOTAL CHARACTERS	80

* Average Length

Exhibit 2-9
Premarket Report
End Uses Item

Data Element Name	Field Type	Field Length
SIC Code	N	9
Description	A	100
Annual Production	N	10
TOTAL CHARACTERS		119

End-use items contain the information about the manner of use intended for each chemical substance. That is to say that end-uses exist only in connection with a particular chemical substance. Again, this information is not likely to change from one annual report to the next.

By-products are very similar to end-uses in that they exist only in connection with a particular chemical substance. By-products may be chemical substances or products. There may be several by-products reported for a chemical substance, and this information may be assumed to remain unchanged over several annual reports.

Production items contain seasonal and annual production figures for a specific chemical substance. It should be noted, however, that production figures are also required for each end-use and each by-product reported. It is anticipated that production figures will change with each annual report, and that all annual reports will be kept on file for checking of trend information.

Exhibit 2 tabulates descriptions of the data elements associated with each item for annual and premarket reports.

2. Data and Report Volumes

In the following discussion of data volumes, the parameters used are estimates representing the best information available from the Office of Toxic Substances. These parameters are presented in Table 1.

Currently, it is anticipated that 3 to 4 reporting regulations and 3 to 4 restriction regulations will be issued each year for the years 1975 to 1980, inclusive. These regulations will cover approximately 500 discrete chemical substances each year for the years 1975, 1976, and 1977. In the years 1978, 1979, and 1980, the regulations are anticipated to cover 1000 discrete chemical substances annually. Each chemical substance will have an average of 10 manufacturers and 10 processors producing approximately 20 distinct reports for each substance. It is estimated that for each chemical substance approximately 25 distinct uses and 4 by-products will be reported. These estimates provide basic implications for the remainder of the study.

Simple data volumes are presented in Table 2. Each row is identified by a number in the left hand column that serves as a key to the following discussion.

One of the major data elements of the OTS Data Base is the chemical substance. Row 5 of Table 2 shows that by 1980, approximately 4,500 distinct chemical substances will be reported to the Office of Toxic Substances by manufacturers, importers, and processors.

Parameters used in Estimating Report and File data volumes

Chemical substances under regulation:

Year	New	Total
1975	500	500
1976	500	1000
1977	500	1500
1978	1000	2500
1979	1000	3500
1980	1000	4500

Average number of respondents for each chemical substance:	20
Average number of by-products for each chemical substance:	4
Average number of end-uses for each chemical substance:	25

Table 1

Annual Reporting Elements

	1975		1976		1977		1978		1979		1980	
	New	Total	New	Total	New	Total	New	Total	New	Total	New	Total
1. Regulations												
2. Reporting	3-4		3-4		3-4		3-4		3-4		3-4	
3. Restriction	3-4		3-4		3-4		3-4		3-4		3-4	
4. Total	6-8	6-8	6-8	12-16	6-8	18-24	6-8	24-32	6-8	30-40	6-8	36-48
5. Chemical Substances	500	500	500	1000	500	1500	1000	2500	1000	3500	1000	4500
6. By-products	2000	2000	2000	4000	2000	6000	4000	10,000	4000	14,000	4000	18,000
7. End-uses	12,500	12,500	12,500	25,000	12,500	37,500	25,000	62,500	25,000	87,500	25,000	112,500
8. Production Reports	10,000	10,000	10,000	20,000	10,000	30,000	20,000	50,000	20,000	70,000	20,000	90,000

Table 2

Another major data element is the end use reported for each chemical substance. Each chemical substance is estimated to have about 25 distinct end uses reported. The volume of end uses reported will grow from an initial 12,500 in 1975 to 112,500 end uses by 1980. The impact of row 7 is to show the number of posting actions that will be required to place the end use information onto separate catalogues for sorting by names of the end uses. This process is similar to indexing a book. The inclusion of the phrase "yellow dye" in the index requires posting of the page number in the index for each use of the word "yellow dye" in the book.

Posting names of by-products is similar to that for end uses. Row 6 shows that the estimated 4 by-products reported for each chemical substance will produce 2000 initial reporting elements which will grow to 18,000 by 1980.

Production figures were estimated at 20 respondents for each chemical substance. Row 8 shows an initial report volume of 10,000 reports which will reach a total 90,000 reports in 1980. These numbers show the growth of the annual report load as the number of regulations in effect grows with the consequent increase in the number of chemical substances under study.

The impact of the premarket reports on the volume of the overall reported data is negligible because the nature of premarket reports is estimated to be such that only one respondent will be reporting a new chemical substance or a new use of an existing chemical substance. Consequently, the impact on OTS will be essentially linear over the six year period while the total number of "new" uses will not increase the end use list significantly.

Table 3, "Estimated Reporting Volumes," shows the impact on the Office of Toxic Substances of the incoming annual reports. For example, in 1980 the Office of Toxic Substances must be geared to handle the reporting requirements resulting from 36 to 48 existing regulations. This means that information about 4500 distinct chemicals will be on file and distributed among 270,000 annual reports. Similarly, 112,500 end-uses will be distributed among the 270,000 reports. This highlights the need for good indexing procedures to provide effective tools for finding desired information.

To draw sharply the distinctions being made, reporting volumes are the number of distinct documents received by the Office of Toxic Substances while the file volumes (Table 4) are the items actually filed for eventual retrieval. The actual files may not contain much of the redundant information reported each year. For instance, information identifying the respondent need not be repeated in the file if no change was reported. Since the reports received constitute legal documents, they must all be preserved in what shall be referred to as the archival files. Whether the archival files will also be the working files is a point for discussion and recommendation. If the archival files and working files are separate files, then additional processes are required with the implied additional

Estimated Reporting Volumes

	1975		1976		1977		1978		1979		1980	
	This year	accum	This year	accum	This year	accum	This year	accum	This year	accum	This year	accum
1. Regulations												
2. Reporting	3-4		3-4		3-4		3-4		3-4			
3. Restrictions	3-4		3-4		3-4		3-4		3-4			
4. Total	6-8	6-8	6-8	12-16	6-8	18-24	6-8	24-32	6-8	30-40	6-8	36-48
5. Chemical Substances	500	500	1000	1500	1500	3000	2500	5500	3500	9000	4500	13,500
7. Product Figures	10,000	10,000	20,000	30,000	30,000	60,000	50,000	110,000	70,000	180,000	90,000	270,000
8. End-uses	12,500		25,000		37,500		62,500		87,500		112,500	
9. Premarket Report	500	500	500	1000	500	1500	500	2000	500	2000	500	3000

Table 3

File Storage Requirements

	1975		1976		1977		1978		1979		1980	
	This year	accum	This year	accum	This year	accum	This year	accum	This year	accum	This year	accum
1. Reports in file	10,500	10,500	20,500	31,000	30,500	61,500	50,500	112,000	70,500	182,500	90,500	273,000
2. Pages		21,000		62,000		123,000		224,000		365,000		546,000
3. Characters*	40,530	40,530	80,670	121,200	120,810	242,010	201,090	443,100	281,370	724,470	361,650	1,086,120

* in 1,000's

Table 4

costs and personnel. A working file separate from the archival file may have a different structure than the structure in which the incoming reports are received. The new structure would reflect the interrogating or query functions anticipated for the file.

The test data file will consist of the actual test data submitted by manufacturers in support of their claims, or at the request of the Administrator. Data items in this file can range from five to six pieces of paper to one or two feet of documents.

Test protocols may also be included in the Test Data File. Anticipated volume of test data is unknown at this time. For estimating purposes, an annual volume of 1,000 reports is assumed for the first 3-year period and 2,000 test data reports are assumed for the second 3-year period.

3. OTS Use of the Data Base

The manufacturer report and test data files will form the nucleus of a data base which will, as it grows, provide for the information needs of organizational units within OTS.

Three of the seven OTS organizational units expect to have requirements on the manufacturer report file: Early Warning, Restrictions, and Testing. The Restrictions Unit will use data reported by manufacturers to justify regulations assist in writing them or assess their impact. The Testing Unit will use the data collected as basis for issuing notices that particular chemicals are being tested. The Early Warning Unit expects to use the premarket reports for issuing warning notices for some substances, as required.

User units will have certain requirements imposed on them, in order to ensure an orderly and useful development of the data base. These requirements include:

(1) identification of chemical substances (or uses thereof) for which test protocols exist; and

(2) identification and collection of existing test data and protocols for inclusion in the OTS file.

As both the data base and OTS experience with it grow, particular attention should be given to the gathering of information on the types of inquiries addressed to the data base by OTS user units. Such user data will provide "profiles" of questions asked by OTS; the profiles can then be used to monitor the responsiveness of the data base to operational needs. The volume of such activity against the data base is also an important factor which measures the work burden placed on the information system and forms a substantial portion of the costs. OTS personnel will also need to monitor administrative activity on the files so that a preliminary estimate of feasible start of automating this activity can be made.

III. FEASIBILITY OF ALTERNATIVE APPROACHES

A. Manual System

1. Information Flow

OTS will publish in the Federal Register a reporting regulation which will list chemical substances subject to study and regulation and will require manufacturers, importers, and processors of the listed substances to report certain information on these substances to OTS.

Upon receipt of a report, a "jacket" is created and a report identification number is assigned for each submission. This ID number is unique for each report and serves as a shelf locator as well as an identifier. In addition to this, the ID number may be used as a "base number" for associating and referencing further submissions or correspondence regarding the report to which it was assigned. The jacket folders are filed numerically by the report ID number; they are then flagged, as appropriate, to indicate when the annual report is due or some other action is required.

Annual reports are reviewed for compliance and correctness. Reports which are incomplete or those which appear to be inconsistent, are forwarded to OTS for verification and/or validation by contacting the reporting establishment. Reports that appear incomplete or otherwise unsatisfactory are verified or corrected by contacting the reporting establishment. A copy of the transmittal letter is filed with the jacket which is flagged accordingly.

Satisfactory reports are forwarded to the cataloging section for indexing and posting of document ID numbers to appropriate catalogs. Six indexing points or descriptors will be required in order to provide an adequate retrieval tool to the file by a logical coordination of the desired terms. These six terms are: (1) names of manufacturers; (2) names of individual chemical substances; (3) names of end-uses; (4) names of by-products; (5) names of new chemical substances; and (6) names of new end-uses of existing chemical substances.

There are many ways of implementing such a manual coordinate indexing system, all of them variants of essentially a single operation: the posting of document ID numbers to the appropriate terms. One of the simplest ways of implementation involves the use of a catalog of 5" X 8" cards with the digits 0 through 9 preprinted across the top of the cards to provide 10 columns. As an example, a report whose ID number is A0432 from the ABC Chemical Company reporting the use of vinyl chloride in the manufacture of paint would be posted in the column indicated by the terminal digit of the ID number, in this case 2, on three cards: one for the manufacturer, the ABC Chemical Company; one for the chemical substance, vinyl chloride; and the third for the end-use, paint.

It is estimated that approximately 32 entries will need to be made in the various catalogs in order to record each submission. Report ID numbers will need to be posted to cards for the manufacturer, for the chemical substance, for an average of 25 end-uses and an (assumed) 4 by-products for each annual report.

Individual test data received will be assigned an appropriate locator number. This number will be recorded on the reporting form submitted by the manufacturer; the test data report itself will be filed by its locator number, apart from the annual (or premarket) report. A test data catalog will need to be maintained in order to provide a search tool for locating desired test data reports, as well as for identifying duplicate test data submissions. It is suggested that the test data catalog be a card file, in test data number order, of short bibliographic descriptions for each test data report. Each test data catalog card will be posted with the ID numbers of annual or premarket reports which submitted the test data. A single, summary card for test data, posted with ID numbers of all submitting reports, will also be necessary for coordination with chemical substance names, as described immediately below.

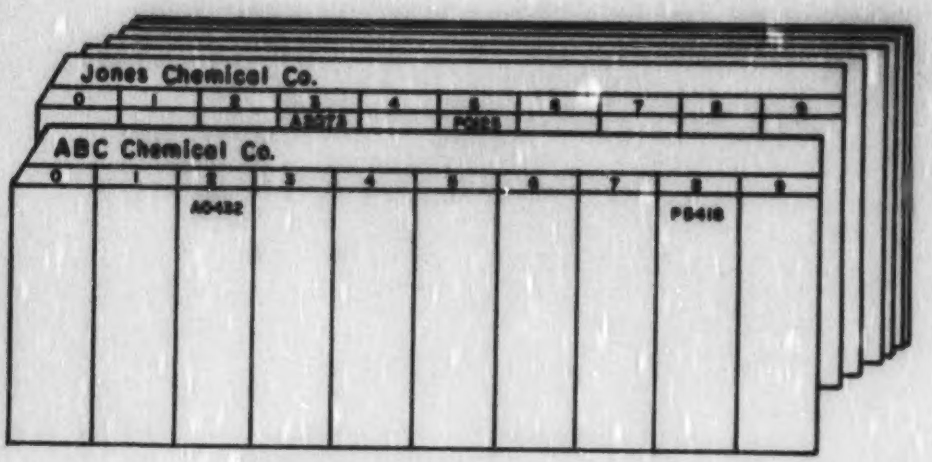
It is very likely that there will be a need for access to the test data file by chemical substance. Such access could be provided by posting the test data ID number to the appropriate card in the chemical substance file. This is not recommended, however, since it is anticipated that the chemical substance cards will be densely posted. A preferable alternative would be a separate card catalog, with cards of a different color, for posting of test data ID numbers. Test protocols could also be accommodated and searched by means of such a test data catalog.

A chronological record will need to be maintained on a current basis.

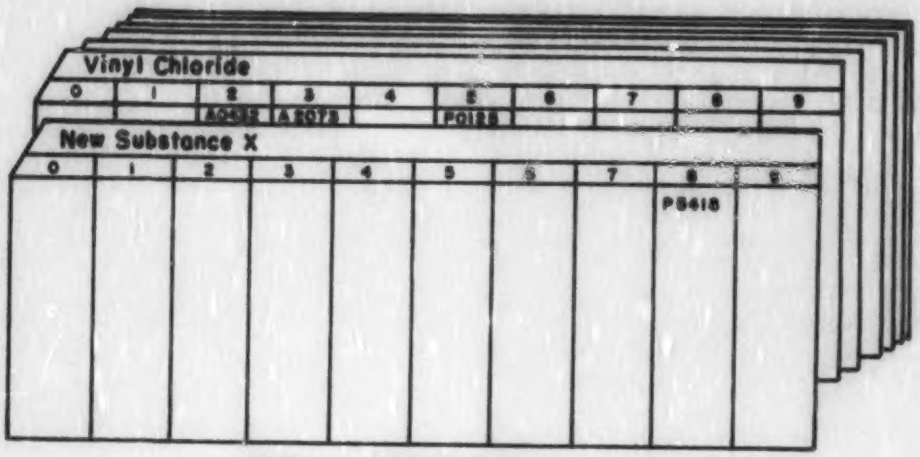
Each entry will contain the ID number of all reports received, plus a short title for each report. For security purposes, the chronological log should not be a card file but rather a book-type record designed to discourage additions or deletions. The purpose of this file is principally to provide full identification of submissions which are recorded in the search files simply as ID numbers.

To minimize the number of catalogs required, it is suggested that new substances be accommodated in the chemical substance file. New substances can be identified by prefixing "New Substance" before the name of the substance (as is done in Figure 2) or by the use of cards of a different color. New uses can be accommodated in a similar manner in the end-use catalog. The by-products catalog will contain names of chemical substances as well as names of products. The distinction can be preserved by the techniques described immediately above. Or, if volume of by-products is extremely low, by-products which are chemical substances can be merged with the chemical substance file, and the remainder with the end-use file.

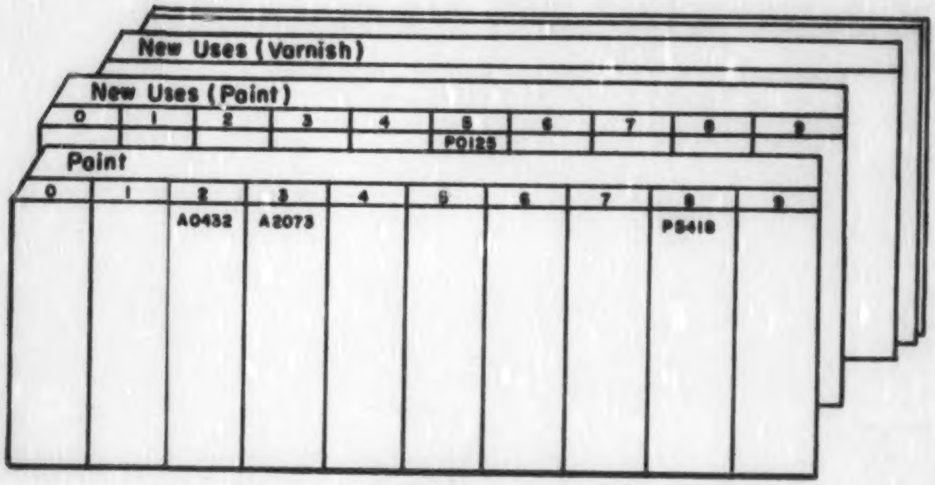
To control access to the catalogs and to safeguard confidentiality of reported information, several techniques are available. One would be to restrict access to catalogs and files to specified personnel only and record all information--both confidential and non-confidential--in the same way. Another technique would be the maintenance of separate catalogs for confidential data. A third technique would be the use of a prefix to the ID number of a submission of indicate that a particular report contains confidential data. The third technique is simplest, and recommended


Jones Chemical Co.
0
1
2
3
A2073
4
5
P0125
6
7
8
9
ABC Chemical Co.
0
1
2
A0432
3
4
5
6
7
8
PS418
9

Manufacturer File


Vinyl Chloride
0
1
2
A0432
3
A2073
4
5
P0125
6
7
8
9
New Substance X
0
1
2
3
4
5
6
7
8
PS418
9

Chemical Substance File


New Uses (Varnish)
New Uses (Paint)
0
1
2
3
4
5
P0125
6
7
8
9
Paint
0
1
2
A0432
3
A2073
4
5
6
7
8
PS418
9

End Use File

Figure 2

for that reason.

ID numbers must be assigned and posted in sequential order. It is suggested that prefixes be used to differentiate between annual and premarket reports, as well as to indicate confidential material. ID numbers can also be suffixed, in order to link any subsequent reports to the original submission, or to reference correspondence regarding a particular report.

2. Retrieval of Information

OTS personnel in need of information from the files would use the catalogs described above by coordinating appropriate catalog cards to retrieve specific documents from the file. To illustrate how this would be done, let us assume that the chronological log of documents contains, among others, the following reports:

P0125 - Premarket Report of Jones Chemical Co., reporting new use of Vinyl Chloride in Paint

A0432 - Annual Report of the ABC Chemical Co., on Vinyl Chloride in Manufacturing Paint

A2073 - Annual Report of Jones Chemical Co., on use of Vinyl Chloride in Paint

P5418 - Premarket Report on New Substance X announced by ABC Chemical Co., for use in Paint

The entries generated in the card catalogs to record the above documents are illustrated in Figure 2. A researcher who needs production figures for vinyl chloride used in paint would pull out the appropriate cards from the chemical substance and end-use catalogs and scan both cards for common entries. He will note that two document numbers are common to both cards: A0432 and A2073. Since both are annual reports they would presumably contain production figures. The researcher then pulls the reports from the files, totals the production figures, and returns the reports to the file. In the course of the search he may note the P5418 entry on the "paint" card. Since the prefix indicates a Premarket Report, the researcher may go to the chronological log of incoming documents to get the identity of the report in order to determine whether or not the report may be of interest to this particular query.

3. Personnel Considerations

In the Manual System, facility personnel will provide essentially library-type services from a library consisting of manufacturer reports, premarket reports, and test data reports. These services can be broadly classified into three types: (1) accessioning and cataloging; (2) information retrieval; and (3) file maintenance.

Accessioning and cataloging personnel will be responsible for the processing of incoming documents and reports. The processing will include functions such as the following:

(1) preparation of jackets for incoming reports;

(2) assignment of proper accession numbers to manufacturer and test data reports to preserve identity and completeness of manufacturer submissions;

(3) maintenance of chronological record of receipts;

(4) editing of data reported, for example, translation of SIC codes to end-use names; checking and correcting spelling of names of manufacturers, chemical substances, etc.;

(5) posting of ID number to appropriate descriptor catalogs;

(6) validation of data reported, including identification of

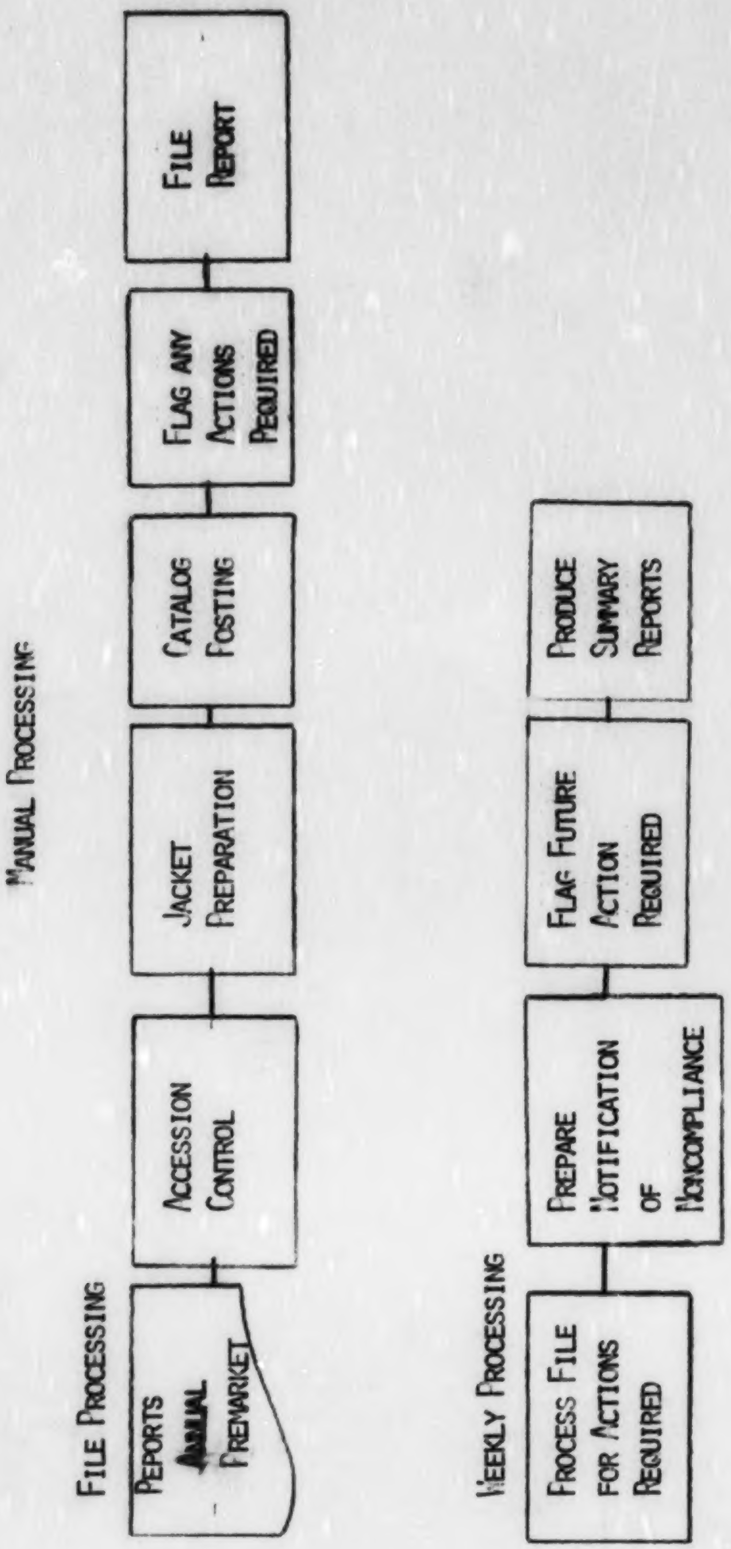

MANUAL PROCESSING
FILE PROCESSING
REPORTS ANNUAL PREMARKET
ACCESSION CONTROL
JACKET PREPARATION
CATALOG POSTING
FLAG ANY ACTIONS REQUIRED
FILE REPORT
WEEKLY PROCESSING
PROCESS FILE FOR ACTIONS REQUIRED
PREPARE NOTIFICATION OF NONCOMPLIANCE
FLAG FUTURE ACTION REQUIRED
PRODUCE SUMMARY REPORTS

Figure 3. Manual Processing

reports for which contact with reporting establishment is necessary;

(7) preparation of short bibliographic citations for test data reports;

(8) maintenance of all catalogs on a current basis;

(9) maintenance of auxiliary catalogs, for example, "next number" record.

To perform the accessioning and cataloging services, three types of personnel will be required: Technical Information Specialist, Data Technician, and Library Technician. The Technical Information Specialist will need sufficient background in chemistry in order to direct and provide guidance to operations in this Section. The Technical Information Specialist will also need to be familiar with the total content and organization of the data base, as well as with the organization and structure of the various catalogs. The Data Technician will review and edit all incoming reports for correct spelling of all entries, for the validity and proper range of production figures reported, and for the forwarding of invalid submissions to appropriate OTS personnel. The Library Technician will be responsible for the preparation, posting, and maintenance of the various card catalogs. The Library Technician will also be responsible for accessioning of incoming reports and for the preparation of jackets for each.

In order to estimate the number of each of the above types of personnel which will be required to perform the necessary services, the following assumptions are made. Each submission will require 40 entries or postings to various catalog cards or logs, as described in the Information Section of this report. In that Section, approximately 32 such entries were identified; this figure is rounded upward to the next tenth for easy estimating purposes, also in order to provide for additional catalogs which may become desirable once operations are underway.

Time estimates for performing these various functions are presented below.

	Minutes
(a) accessioning (entry in chronological log, assignment of ID number)	4
(b) preparation of jacket	3
(c) preparation of catalog card (typing of indexing point to card from appropriate catalog)	1
(d) posting of indexing terms (remove proper card from catalog, post ID number, refile card in catalog)	10
(e) preparation of bibliographic citation for test data reports	5
(f) editing/validation of reports	5

It should be noted here that the above time estimates are intuitive and not supported by hard data. Literature on this subject is generally unavailable.

The chronological log will contain one entry per each document received; test data report volumes are estimated at 1,000 per year for the first three years and 2,000 per year for the next three years. The card catalogs *may contain approximately 400 different manufacturers, 4,500 different chemical substances, an estimated 50,000 different names of end-uses, and an equal number of different by-products. Estimated numbers of postings to each card catalog are presented in Table 3.*

Personnel in the information retrieval section will be responsible for servicing all requests addressed to the facility for any information stored in the data base. These services include the following:

(1) *retrieval of specific documents (manufacturer, premarket, or test data reports) from the files;*

(2) *use of the coordinate indexes to identify documents with specific characteristics, for example, all reports of manufacturers of a particular chemical substance;*

(3) *totaling of production figures from the documents retrieved in (2);*

(4) *preparation of mailing lists;*

(5) *reception of visitors and servicing their requests;*

(6) *flagging of documents to indicate action-required items;*

(7) *maintenance of tickler files;*

(8) *periodic alerting of OTS personnel that preparation by OTS of a particular report or letter is necessary, or publication in the Federal Register is required;*

(9) *accessioning and referencing of all correspondence required.*

Two types of personnel will be required to perform the above service: Technical Information Specialist will direct and provide guidance to information retrieval activities and perform some of the more complicated searches of the catalogs. Record clerks will be responsible for the remaining services specified above.

The following assumptions are made in order to estimate the number of people that will be required for this Section.

	Minutes
(a) retrieval of specific document from file	2
(b) preparation of charge record	2
(c) servicing of queries based on coordination of required terms (analyzing question, structuring query, coordination of appropriate term cards, listing ID numbers)	15
(d) maintenance of tickler files, flagging of items, issuing notices of required action	5

The volume of each of the above functions is unknown at this time. For estimating purposes, it is assumed that one-half of any given year's accumulation will require services (a), (b), and (d), and that approximately 50 questions will be serviced daily.

Duties of file maintenance personnel are self-explanatory. These people will be responsible primarily for the return to the proper shelves of all material used during a given period, and for ensuring that the file(s) returned to the shelves are complete. This section will also be responsible for the maintenance of the log of who had access to which records.

Supervisor and File Clerk personnel types will staff this Section. The filing unit is assumed to be a complete manufacturer report (all 4 items, as presented in Exhibit 2), either annual or premarket, submitted by a manufacturer for one chemical substance. Each report will be filed in its own jacket which will be labeled with the report ID number. The jacket label will serve as a shelf locator. In the case of reports submitted by a single manufacturer for multiple chemical substances, each report for a particular chemical substance will be a filing unit with a different ID number. The jacket for a multiple submission could, but need not be, labeled with the range of numbers included in the jacket. Since report numbers will be assigned sequentially, the multiple submission will still be filed together, either in a common jacket or in individual jackets. Individual jackets do mean increased supplies costs; however, they offer advantages in preserving file integrity and in proper filing of the collection. Individual jackets are assumed here for estimating purposes.

To estimate personnel needed, it is assumed that one-half of each year's accumulation of reports will need to be removed from the files in response to a request and then re-filed after the documents will have served their purpose. The charge records for the documents will then need to be disposed of in accordance with whatever procedure is established. Time estimated per document is 4 minutes. No account is made of the possible overlap between file maintenance and retrieval activities, though it is obvious that the file maintenance section will, more often than not, be re-filing documents.

B. Enhancements to the Manual System

1. Description

This intermediate system is a series of independent progressive steps to enhance the manual system and to move towards a computerized system in a controlled sequence. The steps are arranged in two broad categories and are implicitly batch processes with respect to the Office of Toxic Substances data base. The need to batch process falls from the low volume of input data anticipated for the six year period. This period provides an opportunity to perfect the processing procedures before creating the computerized data base.

a. Automation of the Clerical Tasks Associated with the Indexing of Data Base Documentation

- ° Use of automated data entry systems to facilitate indexing.
- ° Automation of catalog posting and publication.
- ° Construction and sorting of a vocabulary of data descriptors.
- ° Development of a comprehensive Data Element Dictionary across all OTS operation units and in cooperation with other EPA offices.
- ° Preparation of selected and sorted bibliographies.

b. Mechanization of Administrative Clerical Tasks

- ° Automation of mailing list and tickler file for action-required correspondence.
- ° Use of Microfilm as a storage medium for working files.
- ° Development of simple information retrieval through a computerized microfilm retrieval system.

It should be noted that these steps do not require reducing the annual reports or test data reports to machine-readable form and constructing a data base from them. Keypunching reports marks the Computerized System. The steps described above require the accumulation of machine-readable input over a period of time and then performing relatively straight forward sorting and listing functions. The period of performance could be monthly *or even quarterly as need and costs determine. It should be noted, too, that the archival paper files are outside these processes and thus their integrity, completeness, and security are safeguarded.*

2. Indexing Aids

Several existing systems ranging from mini-computers to self-standing data entry systems permit the user to compose his data on the face of a CRT and then preserve this data on machine-readable media. The storage media range from "floppy disc" and cassette tapes to standard computer magnetic tapes and disc units. Essentially these systems give the user great flexibility in correcting and formatting his ideas as compared to a typewriter. Further, the systems can initiate any input process with a structured "form" projected on the screen to prompt the user to provide a complete and more accurate input.

Beyond 1980 the system would first call for the information necessary to prepare a standard OTS bibliographic citation, then the system would permit the abstractor to compose an abstract which will be stored with the citation. Finally, the abstractor would be required to insert one or more descriptors which will be used to index and catalog the document. At the outset the abstractor may only index and catalog.

3. Catalog Posting and Publication

In a manual system the indexing and posting onto several catalogs is a tedious and timeconsuming task with an inevitable chance of error. The costs remain a fairly constant function of the number of additions to the data base. The addition of a new indexing point and, consequently, a new catalog represents a major cost increment. A catalog need not be a card file, however, and several libraries have replaced card files with bound computer listings. After the indexer has assigned values for the various indexing points, the values can be machine entered, rearranged, sorted, and listed to form the several catalogs. The programming is straight forward and operational costs represent a few hundred dollars of processing time.

4. *Data Descriptor Vocabulary*

A means to control the use of descriptors in the indexing process is the development and definition of a standard vocabulary of data descriptors. The standard vocabulary is published and all indexers are constrained to use the terms in the vocabulary.

5. *Data Element Dictionary*

Language (and its use) presents one of the major obstacles to the successful development of data base systems. Wherever possible the OTS should ensure that all its components are using the same terms and in the same ways. Experience indicates that the most successful way to achieve this objective lies in carefully prepared and maintained vocabulary control procedures.

A data element dictionary provides a *specific name* for a datum that the data base will eventually contain and a written, universally available definition of that datum. For example,

COMPANY-NAME	From line 1 of the annual report. This field contains the official and legal name of the company to which this data applies.

CONTACT-POINT	From line 12 of the annual report. This field contains the name and mailing address of the person who is the official contact point between the OTS and the company identified in this record.

The above example shows the manner in which a field receives a unique identifier and how the definition describes the data that will be in that field. In a data base as large as that anticipated by the OTS, the Data Element Dictionary will reach a size that justifies using computer techniques to maintain, update and report its content. Such a file is usually called the Data Directory.

The Data Directory file also has a structure that can be described. Description of all the information needed for the Data Directory about each datum in the OTS Data Base will provide a useful technique for collecting the information and an important management tool for controlling the OTS Data Base.

For management control, the Data Directory may contain the names of the organizational unit responsible for maintaining the accuracy of each datum or for requesting its inclusion in the data base.

6. Bibliography Preparation

Simple bibliographies of test data documents can be prepared from computerized lists of bibliographic citations and indexing points or terms. By sorting and/or selecting on keyword-in-title or indexing terms, lists of similar documents can be produced.

7. Automated Mailing List and Tickler File

The mailing list of respondents and the action-required information can be combined in a single periodic computer run to provide a management report that can highlight potential problem areas and can assist in planning.

8. Microfilm Working Files

The creation and use of microfilm working files lessens the threat to the integrity of the archival paper files and introduces a degree of redundancy that permits the reconstruction of the OTS data base in case of catastrophe.

9. Microfilm Retrieval Systems

Several existing microfilm retrieval systems permit topical searches either through information coded on the microfilm or through computer-assisted indexes stored in computer memory. The latter class of computer-assisted microfilm systems generally uses minicomputers but can also be part of a time-shared environment with large computer systems and distributed terminals. The number of indexing points encoded is limited and access times are constrained by microfilm format and mechanical limitations.

10. Personnel Considerations

The tasks described in this Section are separate and independent functions which can be satisfied by normal service functions. Therefore, no specific skills need be hired to accomplish the tasks but the service organization working on the tasks will require skills as described in the discussion of the computerized system.

However, support of some of the systems on a continuing basis after the initial development will require (or at least be improved by) experience in toxicological indexing and abstracting. These skills may be acquired and kept within the OTS. The indexing keypunchers are an example of such skills.

To avoid the recurring problem associated with introducing automation to existing functions, OTS would do well to seek employee's posessing manual skills who are also familiar with automation and its impact. Successful development of the computerized enhancements to the manual system will be very dependent on the support and cooperation of the existing personnel.

C. The Computerized System

Basically, all industrial reports are processed in a similar fashion. However, the periodicity chosen for the process in each case will reflect the volume of input for each report type. Annual reports with the largest input volume may require daily processing while premarket reports may require weekly or even monthly processing.

Each period's collection of computer input is "batched" into one "input transactions" file and applied against the OTS data base in a serial fashion. As the transactions are applied, prestored programs can perform such tasks as:

- validity check of transactions;
- validity check of the effect of the transaction on the data base;
- checking and reporting any aberrant or noteworthy results from the update (e.g., new totals exceed defined acceptable norms);
- providing administrative reports on actions required by OTS or manufacturers;
- preparation of mailing lists.

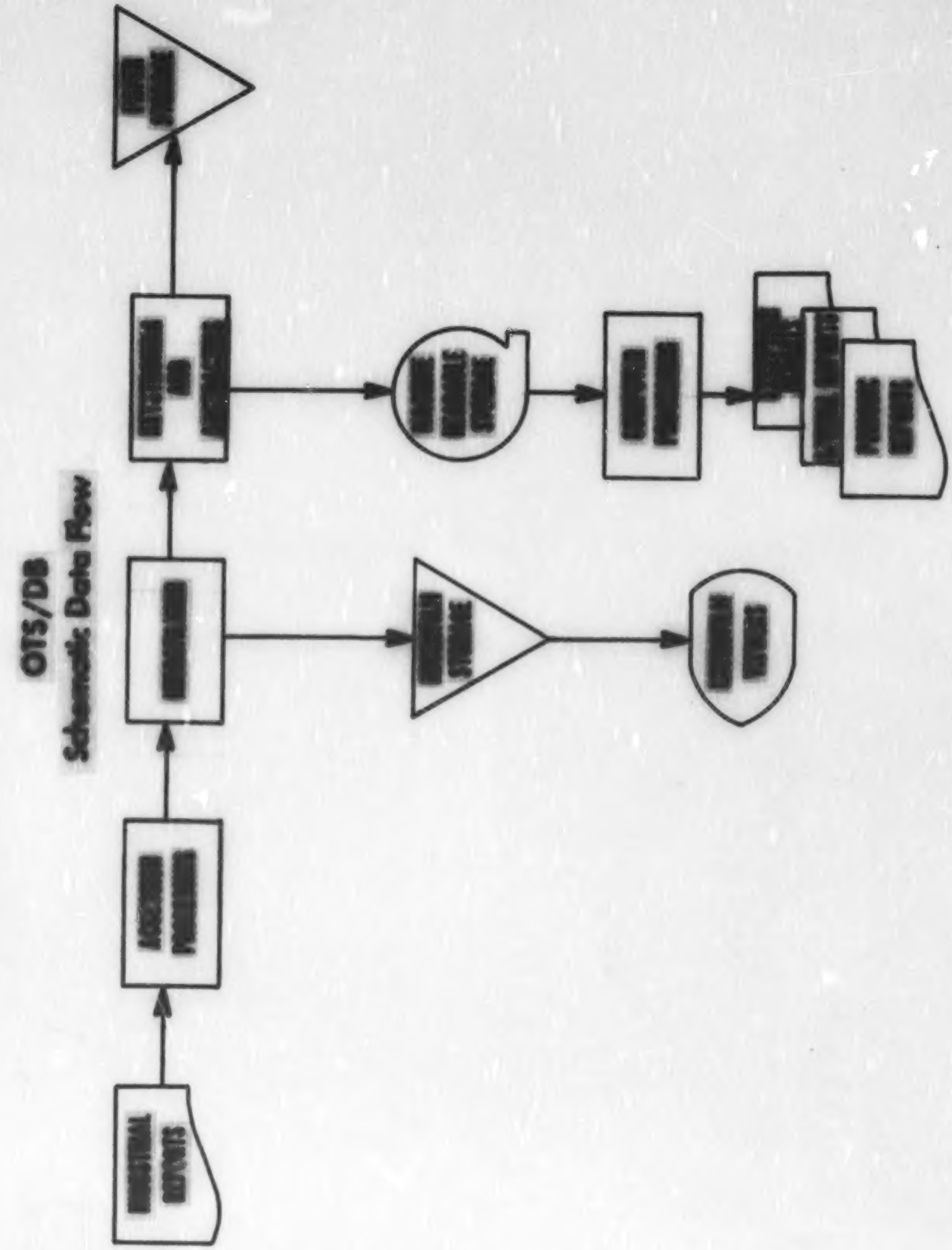

OTS/DB
Schematic Data Flow
INDUSTRIAL REPORTS
ACCESSION PROCESSING

Figure 4. OTS/DB Schematic Data Flow

At the end of the periodic processing, the computer can provide administrative information for planning purposes in the form of summary reports or audit trails.

Queries can be applied against the data base in two general forms. Inquiries which are frequently requested can be prestored as "canned" routines. These can be used periodically to provide the kinds of reports required by the operating units. Other special reports will require specific coding tasks to provide the necessary responses but this process will be materially eased by the data management system with its precisely structured data and available software components. Most systems will provide results in one or two days.

1. Information Flow

a. Mailroom

Because of the confidential nature of many of the reports, special precautions will begin with the arrival of registered mail and a system of accountability traced from the first person signing for the registered mail. The receipted document will be given a control (or accession) number which will uniquely identify the document and be associated with it throughout its existence within the data base. As the number of reports and the number of people using the reports increase, the need to formalize the control of access to the reports will increase, also. This may lead to the first of several administrative files to permit the ready location of a certain document or to reveal who has had access to that document.

The mailroom will also provide the first "sort" of the incoming material into the various categories. This discussion is based on the flow of annual reports.

b. Microfilming

The previously assigned control number will be used to identify the microfilmed report and will become a part of the microfilm record, the original document, and subsequest machine-readable records. The microfilm copy will become the working file and will be available to users in various forms as the needs of the users and the size of the data base evolve. The form of the microfilm storage will determine the manner in which the machine-readable records direct the user to a particular microfilm. For example, if the data is stored on reels of microfilm, the computer record could identify the reel number and position on the reel.

c. Data Entry

After microfilming, the original documents will be forwarded to the data entry section where certain data from the document will be converted to a structured, machine-readable form for entry into the computer for processing.

Proper design of the input document will reduce decisions made at this level to the absolute minimum. Keystroking, in any of its several forms, should be a fairly mechanical task. Several of the newer forms of keystroking permit the operator to record directly on magnetic tape and present the input on a CRT screen so that corrections are immediately and easily made.

Data validation highlights the importance of reducing the keystroking process to a mechanical task. The most common way to insure that the data input to the computer correctly represents the original document is to keystroke the data twice and then compare the two records. In most systems the operator is notified immediately of a discrepancy and is forced to correct it. Data validation has a profound effect on reducing the amount of data errors in the data base.

Keystroking reduces the input into a set of machine-readable characters acceptable to the target computer. This set is usually restrictive in the sense that many characters normally found in the original documents are not available in the machine. For example, the absence of Greek letters will constrain the description of certain chemicals. Standard rules for recovery from such difficulties should be prepared for use by keystroking operators. (For example, using, "alpha", "beta" instead of their symbols.) The set of characters chosen should not unnecessarily limit the ability of OTS to change vendors. Therefore, the chosen character set should be one that is widely available on many computers.

Certain data which may appear on the original reporting document are not readily reducible to machine-readable form. This data includes graphics, mass spectrograms, chemical structure figures, charts, and drawings. Additionally, cost effective considerations will argue against inputting to the computer lengthy narrative descriptive material or even abstracts of such material for full text searching.

d. Encoding, Abstracting, and Quality Control of Data

Encoding is the reduction of a concept, whether simple or complex, to a single datum, either a word, alphanumeric coded character string, or numeric value. Abstracting is the reduction of a lengthy narrative description into a succinct equivalent. Both processes ought to reduce the size of the explanation without losing any information or meaning. Quality control of the data base is the continuous process of assuring that the data base is complete, accurate, and current.

Computers can check for clerical errors such as missing data, erroneous or nonsensical codes, and numbers outside valid ranges but cannot check miscoding or make correlations that human checkers can. The quality of the data base should be the specific responsibility of a quality control unit that will make decisions and use all available resources to insure the highest integrity and completeness of the data base.

Clearly, maintaining the integrity of the data base requires a resource expenditure which increases as the requirement for accuracy increases; the more spent, the better the integrity. Indirectly, the data base is quality checked by the users who point out inaccuracies and incompleteness. But it should be noted that increased usage justifies increased quality control -- often as a result of user screams.

Certain portions of the annual reports will require human intervention in order to understand and abstract the reports being submitted. For example, the Bureau of Census has a very structured reporting form reflecting their wealth of experience in collecting and building data bases of information, but they have found the need to use a category coded "other" for the inevitable surprises that respondents produce. The "other" code requires a technician to analyze and record the unexpected answer. Furthermore, the use of "other" codes should be monitored in order to provide for new fields or changes in existing ones, as soon as such needs arise.

2. Data Entry Considerations

Data entry, the point where information enters the system, represents a highly crucial point in the flow of information from the manufacturers of toxic substances to the OTS. Data entry is the largest cost element in any machine-oriented system. It is prone to exaggerated errors because of the cascading misinformation resulting from bad input. Bad design in the data entry aspects of a computer system foredooms the entire system. This Section discusses various methods of data entry in order to review options without a close scrutiny of the feasibility or economics of the methods for the OTS application.

To place data entry in context, data entry begins at the point where OTS acknowledges receipt of incoming data. At some point, some agent of OTS must decide whether the data falls within the scope of the OTS data base and should, therefore, be included. Data entry ends with the insertion of all pertinent data within one or more storage units of the system.

In terms of design, data entry costs should help determine the eventual

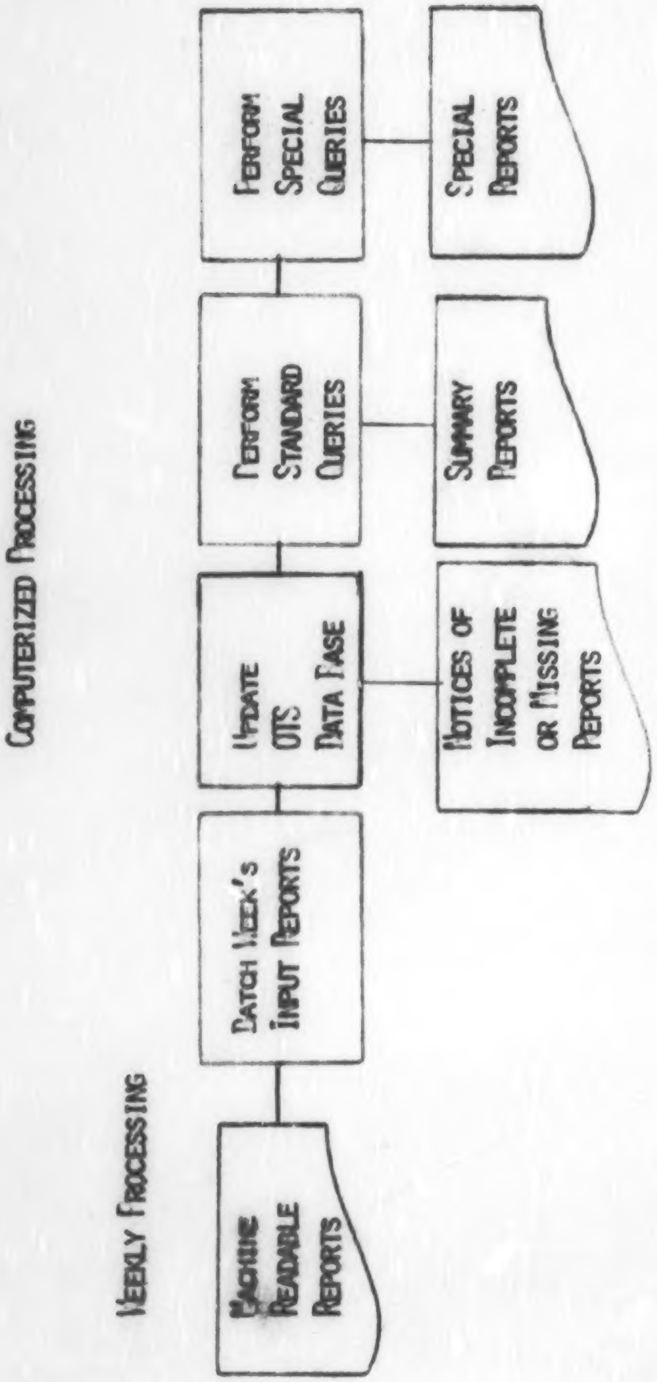

Computerized Processing
Weekly Processing
Machine Readable Reports
Batch Week's Input Reports
Update OTS Data Base
Notices of Incomplete or Missing Reports
Perform Standard Queries
Summary Reports
Perform Special Queries
Special Reports

Figure 5. Computerized Processing

scope and size of the OTS data base as well as the frequency of input processing. Similarly, the design of the data file structures, interrelationships, and storage media will affect the design of data entry systems. This means an inevitable iterative approach to designing both.

From a procedural standpoint, data entry rests heavily on the quality of the input documents, the training and capabilities of the data entry personnel, and the accession methods chosen in each distinct file. No manager should be surprised that computer system design must reflect the question of personnel availability and that using the computer to replace clerical help often results in higher operating costs. Whether made consciously or not, the decision between faster response and further dependence on clerical staff often ignores the costs involved.

Data entry methods fall into two broad areas: keystroke and optical character recognition, (OCR).

Keystroke methods depend on the human eye, brain, and fingers. All data flow from the document to the storage medium, therefore, rests on the inherent weaknesses and strengths of people. Examples of keystroke methods are:

- Card punching. This classic method of data entry places data on 80-column punch cards. This method is characterized by many inherent limits, such as difficulty in correcting errors, ease of misarrangement of data, large storage requirements, very slow input rates and consequent high costs. But this method is universally used, nevertheless. This last fact, always surprising, means the ready availability of personnel, job shops for peak loads, and competitive pricing. (Some shops now use Korean and Taiwanian help to cut costs.)

- Key-to-tape. The keystrokes record directly on a magnetic tape and result in much larger records, ready error correction, reduced media and computer entry costs, and very reduced storage requirements. Though usage is growing rapidly, key-to-tape systems are several orders of magnitude less available than card punching systems.

- Key-to-drum. Though quite similar to the key-to-tape method in that the keystrokes record directly on a magnetic storage medium, the random storage structure of the drum permits the "clustering" of many input terminals around one drum. The necessary control for such clusters of one device is often a minicomputer.

- Intelligent terminal. This method resembles the key-to-drum above except that the control computer can be used to structure the input in an interactive dialogue with the terminal operator. In effect, the terminal asks questions, displays the proper format, and indicates detected errors. Additionally, the configuration of the intelligent terminal may contain communication links for direct transmission of data to the main system and its storage facilities.

- Re-typing in OCR font or format. This is a special form of keystroking or mark-sensing which reformats documents not originally prepared in a form acceptable to OCR. Its main advantage over other forms of keystroking lies in the output which both humans and machines can read.

Mark-sensing methods permit a machine to read directly a previously prepared document. Mark-sensing includes such options as sensing special pencil marks, Magnetic Ink Character Recognition (MICR), point-of-scale data recognition devices, etc. However, this report considers only optical character recognition.

OCR permits the scanning of printed material directly by mechanical devices to translate the data into machine-readable form with little or no human intervention. One group of OCR devices uses microfilm as input.

A major disadvantage of OCR lies in the initial capital expenses of acquiring this fairly sophisticated equipment. However, current technology indicates significant economies for unit processing costs. For OTS this means using vendors who have absorbed the initial costs or who have amortized them over a period of time. Such vendors do exist in the Washington, D. C. area.

OCR offers particular advantages when the user has control over the format and structure of the input material. Further, the process permits scanning only a portion of the material first and performing a more complete scan at a later time.

3. Data Management System Selection

The data management system selected to provide the major portion of the processing should be a proprietary package available to all potential vendors and to ADP facilities within EPA or other important government agencies. By selecting a proprietary package from a reputable supplier, OTS is reasonably assured that the critical kernel of its information handling capability will have continuing maintenance and support. Further, certain proprietary data management systems have implementations on several vendors' equipment and offer some degree of flexibility in selecting contractors.

As a guide to selecting one, from among the many available, data management system to serve the OTS data base note the following important considerations:

- Interface with a common procedural computer language;
- Flexible report writer capability;

- *Basic orientation to sequential file processing but extensible to random accessing;*
- *System support of variable length files and fields, with grouping in a hierarchical data structure;*
- *Extensible to remote processing;*
- *Audit trails of data processing activities;*
- *Extensive data validation features;*
- *Ease of file loading (i.e., inputting the initial values to a newly defined file);*
- *Flexible repertoire of file commands;*
- *Output control with selective language facilities which permit constraining the output to pertinent areas of the files.*

4. *Data Management System Review*

Of the approximately 40 data management systems reviewed, three were selected as candidate systems with the necessary capabilities and features required by OTS data processing needs. These candidate systems are GIS, RAMIS, and System 2000; summary descriptions of each are presented separately at the end of this Section.

Common to all of these systems are the capabilities to perform the functions basic to all data base management systems. Each has a data description language which enables the user to define both logical and physical characteristics of the OTS data base. Each has a capability to coordinate and supervise the hardware and software components necessary to process the data. Each provides the capability for all users to process required data without concern for its physical location or for its logical relationships with various user application programs.

In addition to the capabilities common to all data management systems, the candidate system must have other capabilities to satisfy requirements peculiar to OTS. These capabilities are: (1) computational facilities which include arithmetic, logical, and relational operations; (2) hospitality to user application programs necessary to provide other computational services not included in the candidate system but required by OTS; (3) multi-file processing; (4) sorting of records in user-specified order; (5) flexible, user-specified reporting formats; (6) hospitality to a hierarchical data structure, with variable length records, permitting multiple occurrences of fields at all hierarchical levels, (7) audit trails; and (8) data validation facilities.

The above specifications are based on current understanding of OTS requirements. These requirements are reviewed below, with discussions keyed to the numbers in the preceding paragraph, as well as to paragraph numbers in the summary descriptions.

(1) Computational facilities provided with many data management packages include arithmetic operations (add, subtract, multiply, divide); Boolean operations (and, not, or, and combinations of these); and relational operations (equal to, less than, greater than, etc. and combinations of these). OTS has indicated a need for all of these capabilities; and all candidate systems have them.

(2) Hospitality to user application programs is necessary to supplement computational facilities not provided by the data management system in order to enable the use of existing (or proposed) OTS application programs with the OTS data base. To date, one application program has

been identified - an EPA program which relates CAS numbers to names and synonyms of chemical substances. All candidate systems can accommodate application programs written in procedure-oriented languages.

(3) Multi-file processing, or the capability to open more than one file at a time is necessary principally for efficient file management. Two files have been identified as probable initial components of the OTS data base. These files comprise records of disparate structures and lengths; individual records within files may have disparate rates of activity. For example, address portions of annual reports will need to be retrieved much less frequently than annual production figures for a given chemical substance. There may be other such records whose volume of activity or similarity of structure will warrant file segmentation in order to optimize overall storage and processing efficiency.

(4) Sorting of records in user-specified order is a requirement included principally to provide some control over the information retrieved from the data base in response to a query. As an example, information satisfying search criteria in a particular query may be output in the order in which the source data was entered, or in some order peculiar to the data management system in question. This may not always be the best way to present the information to the user who may need mailing lists, for example, arranged alphabetically by manufacturer, or annual production figures for a chemical substance listed in descending order of magnitudes.

(5) Flexible, user-specified formats will provide additional latitude in specifying print or display formats for computer output. This includes options such as user-supplied report titles or column headings,

tabular output, etc. For very complicated reports, there is the additional capability provided for storage and subsequent retrieval at processing time of frequently used routines for complicated search requests or report formats.

(6) Hospitality to a hierarchical data structure accommodates multiple occurrences of variable-length records at all levels. The OTS data base is such that relationships among data elements need to be identified and preserved, and fields at all hierarchical levels must be searchable in interrogating the files.

(7) Audit trails keep records of changes made to files, of access to various data fields, and provide similar information necessary to monitoring data base usage. In case of destruction, audit trails together with backup copy can be used for reconstruction of files. Audit trails are important also in safeguarding confidentiality of data. In this respect, the three candidate systems offer password control which may be used to restrict access to individual fields.

(8) Data validation on input includes range checking, checking for alphabetic characters in numeric fields, etc. The candidate systems have some capabilities for data editing.

As a selection tool, comparative feature analysis of data management systems has its limitations, especially because of the fact that there are about as many ways to implement features as there are features. More knowledge of the data would aid in providing criteria for selection of the data management system most suited to OTS needs.

More valid estimates of the system could be prepared during the initial phase of operations of the OTS facility. Specifically, a machine file of the first 1000 manufacturer reports could be constructed and used as an experimental data base for testing the capabilities of System 2000 which is available at EPA. Performance data gathered for the experimental data base could then be used as a benchmark in the eventual procurement of a data management system, should another or different data management system become necessary at some time in the future.

During the initial phase of operations OTS use of other data bases will need to be explored. Currently, OTS is using (or planning to use) seven data bases: ENVIRON; TOXLINE; WORD-ONE; NIOSH-BERT; AMIC; PARCS; and OWH. Requirements of OTS from each of these will need to be determined so that appropriate interfaces can be provided between each auxiliary data base and the OTS data management system.

Source documents used in preparing this Section include the following:

(1) Auerbach Computer Technology Reports, Vol. 51, Software Reports, Auerbach Publishers, Inc., Philadelphia, Pa., 1972, and periodic page revisions.

(2) Data Management Systems Catalog, Mitre Corp., Bedford, Mass. Rept. No. MTP-139, January 1973, 394p.

(3) DATAPRO 70. The EDP Buyer's Bible, Datapro Research Corp., Delran, N.J., 1973, and periodic page revisions.

Summary descriptions of the three candidate systems follow.

PACKAGE: GIS (Generalized Information System)

SUPPLIER: IBM Corp., White Plains, N.Y. 10601

COST

Purchase:
Rental : $450 month for Basic Retrieval System, prerequisite for all features

SYSTEM REQUIREMENTS

Computer(s) : IBM 360/70 (with decimal arithmetic feature)

Core Storage : 113K bytes; additional 10K bytes with teleprocessing feature plus 15-30K bytes if online terminals are used

Auxiliary Storage : Minimum of three 2311 disks recommended

Input/Output : Card, printer, disk

Operating System : OS

(1) Computational facilities: "Standard logical operators (e.g., greater than, not equal, between)." Other operators included with system options: increase, decrease, multiply, divide; total; average; unicount (number of unique values of specified fields); scanning for occurrences of specified strings; detection of increases/decreases in field contents since last tested; and detection of fields with blanks or zeroes.

(2) User Program Interface: User-written code must perform file maintenance with the Basic Retrieval System; file modify and update options necessary for GIS to perform file maintenance. The BRS can pass control to user programs written in COBOL, FORTRAN, PL/1, or assembler language and provide the program with data retrieved in a query.

(3) Multi-file Processing: Up to 3 files concurrently; optional feature extends multiple processing to a maximum of 16 files.

(4) Sorting: Alphabetic, ascending, descending on up to 64 keys. Maximum of 115 fields can be extracted from up to 3 files and assigned to maximum of 16 temporary files for sorting and printing independently.

(5) Reports: Optional features control page formating, (size, spacing, labels) subtotals, data suppression, and offline printing capability when operating from online terminals. Utility option provides capability to maintain library of source language programs and precompiled procedures.

(6) Data Structure: Basic system supports non-hierarchical files processed in sequential or indexed-sequential mode. Hierarchic file support option permits repeating groups at up to 15 levels, with only 1 repeating group permitted at any particular level, except the lowest.

(7) Audit Trails: Access control is provided at both file and field levels.

(8) Input Checks: The Edit and Encode option provides for editing of input data which includes range and validity checks based on specifications from the file definition. It also provides for the automatic encoding of source data values. File and field access controls govern authority to retrieve and modify data.

COMMENTS: IBM Class A Program Product. Current applications are in industries requiring querying of large data bases; number of installations not indicated.

PACKAGE: RAMIS (Rapid Access Management Information System)

SUPPLIER: Mathematica, Inc., Princeton, N.J. 08540

COST

Purchase: $21,000
Rental : $500 per month; $3,000 extra for TS version

SYSTEM REQUIREMENTS

Computer(s)	*IBM 360/40 and Up; IBM 370*
Core Storage	*128K*
Auxiliary Storage	*1 disk*
Input/Output	*terminal; tape; card reader; printer; disc*
Operating System(s)	*OS*

(1) Computational Facilities: Mathematical operations permit running calculations to be performed on data in given fields. These operations include average, maximum, minimum, average sum of squares, and percentages. Other operators include arithmetic (plus, minus, divide, multiply); relational (equal to, less than, greater than, not equal to, and combinations of these); functional (minimum, maximum, square foot, exponentiation, log, absolute value, integer part) and logical (and, or, if then). Other facilities offered include concatenation, date conversions, and field editing.

(2) User Program Interface: Host language interface allows programs (COBOL, FORTRAN, AND PL/1) to access (by using field names) and change RAMIS records directly. Data returned from one or more files can be retained

for further processing within RAMIS, saved as sequential records that can be passed to other programs, or incorporated into reports.

(3) Multi-file Processing: Up to 10 files can be linked together for processing as a set.

(4) Sorting: "Various types" (from next paragraph).

(5) Reports: May be produced in standard formats or in formats specified by the user, and may include results of calculations which provide column totals, sub-totals, row totals, percentages; as well as various sorts and edits. An option is an output package that generates histograms and plots with two or more axes.

(6) Data Structure: Hierarchies are established through tree-structured organization of files with cross referencing. Linked lists of fields are maintained and accessed by the system through indexes. A maximum of 24 levels is allowed with repeating groups with variable numbers of occurrences for each group.

(7) Audit Trails: Maintenance of log of rejected records (in file updates).

(8) Input Checks: Edit and audit checks on all files; directory name checked against field names in search requests.

COMMENTS: First installed in 1967; 500 copies now in operation throughout the world. Available through National CSS time-sharing service at $600 per month.

USER COMMENTS: "A well-established package that has been enthusiastically received in the marketplace. Users contacted are extremely happy with system. They say data can be easily obtained when needed by nonprogrammers - including technical, administrative, and clerical personnel. . . . Installation and training support are rated good to very good; ongoing maintenance very good to excellent. Supplier personnel respond rapidly and effectively to all customer problems. Documentation is considered adequate but improving. Minor deficiencies surfaced, most seem unique to a particular user. One user has 100K-byte partitions, and these are too small for RAMIS. Other problems encountered were associated with the system's inability to handle other data files without previously converting to RAMIS format, which means that duplicate files must be maintained when files are used by both RAMIS and other applications. However, a host language interface can be used to tie RAMIS files to the system." (Auerbach)

PACKAGE: System 2000

SUPPLIER: MRI Systems Corp., Austin, Texas 78766

COST

Purchase: $25,000 (paid-up lease, one time charge)
Rental : $740 per month; 960 rental with option

SYSTEM REQUIREMENTS:

Computer(s) : CDC 6000 and Cyber 70 Series; IBM 360/40 and up; IBM 370/135 and up; Univac 1100 Series

Core Storage : 16K decimal words; 90-150K bytes; 22K words

Auxiliary Storage : disk and at least 2 mag tape drives if Sequential File Option is used (for all computer systems)

Input/Output : standard I/O for each computer system

Operating System : Scope, Kronos; OS, DOS; Exec 8

(1) Computational Facilities: The following relationships may be specified: logical (and, or not); comparative (equal, not equal, greater, less, or combinations of these); range (for example, from 30 to 40 lbs); and existence (whether date of a transaction is entered). With the Immediate Access Module, calculations such as tally, sum, average, maximum, and standard deviation can also be specified.

(2) User Program Interface: The Procedural Language Feature enables users to process System 2000 data with application programs written in any part of the data base, retrieve data in desired sequence and format, and update the data base with the application programs.

System 2000 statements are embedded in these programs and, when encountered, they activate System 2000 through the Interface Module.

(3) Multi-file Processing: Can process entire files from tape, or files that are divided between tape and disk. Efficient support is available for data bases that range in size from a few thousand to hundreds of millions of characters.

(4) Sorting: The Immediate Access Module permits users to specify sorting on up to 40 keys.

(5) Reports: The Reporter Writer provides extensive formatting flexibility including column, row, and page headings; dates; footnotes and other explanatory notes; nested control breaks with totals and sub-totals; and ordering of report contents. Reports can also contain computed values whose formats are derived from their field contents and literals. Report formats can be cataloged and stored to avoid repetitive redefinition.

(6) Data Structure: Inverted file structure oriented around the repeating group can theoretically accommodate 32 levels. Practical limit however, is 8 to 10 levels. Up to 430 data elements may be included in one or more repeating groups.

(7) Audit Trails: Creates machine-legible file of updating transactions that can be used with archive copy of data base for audit or backup purposes. Password control for entire files; additional passwords may be used to control access to individual fields.

(8) Input Checks: Source data checks against user's file definition.

COMMENTS: The package was first released in June 1970 and is now (October 1973) installed in more than 40 locations. Also offered through various service bureaus nationally, and internationally through CDC's Cybernet time-sharing network.

USER COMMENTS: "System 2000 users who were contacted feel that it is an outstanding package. They indicate that the package is easily used for all data management functions, emphasizing its flexibility in redefining file formats and its retrieval capabilities that permit nonprogrammers to structure rapidly both batch and on-line requests. The package is used easily and effectively with programs written in other procedural languages.

These users indicate that System 2000 accomplishes its functions rapidly; they attribute this primarily to indexes and other structural pointers in the data base. However, the increase in speed is achieved at the cost of storage space. In this tradeoff, which is almost unavoidable, the users prefer the speed of System 2000.

Support provided with the package is rated very good. Supplier personnel are knowledgeable and cooperative, and they maintain close contact with System 2000 users." (Auerbach)

"Users of System 2000 contacted by Datapro 70 were unanimous in their praise for the product and its vendor. The users stated that System 2000 performs as advertised and yields data base expansion ratios of 1.5:1 or less when the user keeps within the 20% keyed item ratio recommended by MRI. Datapro interviewed users with IBM System 370 and CDC Cyber 70 Systems. One user had particular praise for two features in the latest release he's received: support for 1000 elements (up from 430) and multiple data base support in procedural languages (which allows the user to work with second, third, and fourth data bases while System 2000 automatically keeps track of the user's position in each data base). This user also plans to adopt the multiple-user option of System 2000. All users said that response times for data base inquiries were very fast. In general, System 2000 offers excellent potential for high-speed on-line information retrieval based on complex requests." (Datapro)

5. *System Configuration*

The crucial kernel of the computerized information handling capabilities is an off-the-shelf proprietary data management system selected for use on the OTS data base. In this discussion, the System 2000 Data Management System is used as an example of a typical data management system and is used to provide the parameters for estimating the system configuration necessary to support OTS computerized operations described herein.

According to available data, the System 2000 Data Management System requires the following computer configuration:

Mainframe:	*Univac 1110*
Memory:	*131,000 words*
Peripherals:	*2 Magnetic Tape Handlers* *Univac FASTRAND III*
Teleprocessing:	*Equipment evailable should such an extension become necessary.*

It should be noted that System 2000 can operate on several distinct vendor computer systems but the Univac 1110 is available to EPA and provides a basis for cost estimates.

6. *Personnel Considerations*

In the Computerized System facility personnel will provide the same library-type services from the same library as in the Manual System, with assistance from a computer to automate many tasks as described below.

Accessioning and cataloging personnel will be responsible for the processing of incoming documents and reports and conversion to machine-readable form of all document descriptions and of entire annual and premarket reports. Functions outlined for this Section in the manual system will also need to be performed here, except that manual posting of indexing terms to appropriate catalogs will be unnecessary. In addition to this, maintenance of catalogs will be automated. However, some maintenance activity will be required in regard to catalogs, for there will be a need to maintain a catalog of records in process or ready to be processed in the next scheduled output of a computer-produced catalog.

Additional functions for accessioning and cataloging personnel include:

- *(a) construction and sorting of a vocabulary of data descriptors;*
- *(b) preparation and scheduling of documents for microfilming;*
- *(c) indexing of documents;*
- *(d) development of a comprehensive Data Element Dictionary for all OTS operational units.*

Personnel types required to fulfill accessioning and cataloging functions in the Computerized System include Technical Information Specialist; Data Technician; Analyst/Programmer; Keypunch Operator; and Library Technician. Required competence of Data Technician personnel will remain essentially as that in the Manual System. Library Technician personnel will, in addition to their accessioning responsibilities cited above, have the added responsibility of control and scheduling of documents for microfilming.

Analyst/Programmer personnel will guide and participate in the preparation of the Data Element Dictionary. Keypunch operators will transcribe all necessary records for machine storage. In order to estimate the number of people required, the following assumptions are made.

All data elements reported in the annual and premarket reports will be converted to machine-readable form. Total number of characters needed is 4014 for manufacturer reports, and 780 for premarket reports as described in Exhibit 2.

Time estimates for performing these various functions are presented below.

		Minutes
(a)	*preparation of document for microfilming*	*10*
(b)	*maintenance of controlled vocabulary*	*1.5*
(c)	*data conversion*	*150 keystrokes/ minute*

Information retrieval personnel will be responsible for servicing all requests addressed to the facility for any information stored in the data base. Though the services provided will remain essentially the same as in the Manual System, the tasks associated with information retrieval activities will be automated. Information retrieval personnel will, therefore, use different tools and have three files to work with: the physical store of documents; the machine file of manufacturer report surrogates; and the microfilm store. The services that will be required are the following:

(1) retrieval of specific documents from any of the files;

(2) reception of visitors and servicing their requests;

(3) scheduling computer operations to provide:

  - periodic indexes to holdings
  - vocabulary of data descriptors
  - selective bibliographies
  - Data Element Dictionary updates
  - periodic and aperiodic mailing lists
  - notices of action required by OTS or manufacturers
  - indexes to microfilm stores

(4) preparation of routine queries for interrogating the computer file ("canned" search routines); and

(5) preparation of non-routine queries for retrieval of documents or information from the machine file.

The types of personnel required to service these information retrieval operations include: Technical Information Specialist; Analyst/Programmer; Equipment Operator; and Record Clerk. The Technical Information Specialist will be responsible for structuring technical search questions addressed to the computer. The Analyst/Programmer will convert the questions to appropriate programs. This person will also be responsible for analyzing other problems presented for solution by computer and for preparing the required computer programs. This person will also be responsible for the production and periodic updating of the Data Element Dictionary. Equipment operators will be needed to operate and maintain microfilm readers, printers, computer terminals, and data conversion and communications equipment. Record clerks

IV. EVALUATION AND COSTING CONSIDERATIONS

A. Evaluation Considerations

1. Evaluation Criteria

In comparing the various options available to OTS, it is important to understand the various facilities provided by the options as well as the cost of the optional method. The manual system can be compared in function and cost to the computerized system. The enhanced manual system necessarily contains the manual system but the computerized enhancements may reduce clerical tasks. On the other hand, the computerized approach may provide a capability that is new or even impossible under a purely manual system. An excellent case in point is the number of indexing points available under the various systems.

An indexing point represents a pre-stored answer to an anticipated question. For example, to file all the manufacturer annual reports alphabetically by manufacturer name anticipates the need in the future to find a specific report if one is given the manufacturer's name. To keep a card catalog of chemical substances which are filed under the various annual reports anticipates that someone may seek information about specific chemical substances. A manual system with indexing points requires manual posting of reference points onto the indexing terms represented by the catalog cards. Consequently, each new requirement for an indexing point means a significant increase in the labor of posting new entries but it also means a significant task of retrospectively re-indexing all previously indexed file entries. A computerized system can trivially increase its indexing capability by one indexing term.

The criteria used in comparing the cost/benefits of the three systems are listed below with an explanation of each.

a. Indexing points capacity. The number of search term points available to the user who is seeking information from the data base. Each search point or indexing point may have many possible values. For example, the indexing point, Chemical Substances, may have as many different values as there are distinct chemical substances entered in the data base.

b. Input lag. The amount of time that the data base is behind the actual status of the real world.

c. Response time. The length of time necessary to receive an answer to a query.

d. Data extensibility. The ability to receive new kinds of data.

e. Query extensibility. The ability to answer new kinds of questions.

f. Reporting facilities. Various aids to present any information from the data base.

g. Availability of system components. Whether off-the-shelf or requiring developmental time.

h. Personnel constraints. The requirement for special or hard-to-find skills.

i. Costs.

j. Security. The ability to meet the confidentiality requirements of OTS.

2. Facilities Evaluation

In order to perform the operations envisioned under the various systems, the following facilities are required:

- Administrative processing and clerical support;
- Forms design capability;
- Photocopying and binding;

- *Secure storage facilities;*
- *Performance site;*
- *Indexing.*

Each of these facilities will be discussed in sufficient detail to permit the Office of Toxic Substances to evaluate compliance of offerors of these services.

a. *Administrative Processing and Clerical Support*

In a manual system the need for clerical support is quite obvious. Less obvious and requiring special underscoring is the rather large administrative burden represented in the Toxic Substances Control Act. The Office of Toxic Substances faces a maze of interlinking dialogues with its respondents, both individually and collectively. It can reasonably be expected that many of the respondents will be hostile and uncooperative and will require carefully collected correspondence files to protect the goals of the Office of Toxic Substances. A fairly typical scenario might begin with a query from a manufacturer as to whether his use of a certain substance constitutes a new use. OTS's answer must include justification which must come from its data base. If the new use requires test data, OTS must show that such data is not presently available and must provide the manufacturer with a test protocol, if a protocol is requested. From the time the respondent returns the required test data, OTS must insure that future use of the test data by other respondents will result in their paying a proportionate share of the testing costs. This entire string of events is punctuated by administrative actions to insure compliance, follow-up, valid records as the basis of decisions, legally structured demands, and fair proceedings.

To a large extent, this administrative burden can be satisfied and anticipated with good procedures. It cannot be ignored without jeopardizing OTS's credibility.

b. Forms Design Capability

The correctness and completeness of the OTS data depend to a large extent on the clarity and precision of the data requesting medium. Even without the eventual goal of computerization, OTS must develop standard respondent forms that help to discipline and structure the growing data base. Forms design will become a continuing iterative task as procedures develop and the need for more specialized forms increases.

c. Photocopying and Binding

The very nature of the OTS responsibilities argues for the fact that the physical construction of the many reports and much of the correspondence prevent treating this facility as an aside. The area of dissemination, the volume of data, its economic and social importance, the potential of legal action, all argue for the availability of this capability. The need to present OTS's position in a correct and professional manner requires special attention to this facility.

d. Secure Storage Facilities

OTS must insure that its responsibilities and proprietary rights to the information being collected and generated are being satisfied. The storage facilities must protect OTS's data base from intentional and accidental loss or damage. The problems and considerations in this area are discussed fully in the Confidentiality Section. However, to provide some indication of storage requirements, the following estimates are presented.

For the volume of annual reports estimated to be 10,000 reports in 1975, and assuming one-eighth of an inch for each report, the storage required is 12.5 four-drawer filing cabinets or 38 square feet of floor space. By 1980 there will be 270,000 reports on file. That amount will require 338 four-drawer cabinets or 1000 square feet of floor space. The retention term of the various data base documents is not known.

e. Performance Site

Office of Toxic Substances requirements for quality control and confidentiality lead to certain conclusions about the performance site of the several steps involved in processing the information. To meet its responsibilities, the Office of Toxic Substances must know precisely where the processing steps are executed and what physical elements exist to provide the protection needed.

To provide an outline for the following discussion, the possibility of five discrete sites is assumed but the reader should realize that certain of the steps are electives OTS may not select, and all steps might be performed at one site. The five sites are:

- *Office of Toxic Substances;*
- *Mail receipt and accession control point;*
- *Microfilming;*
- *Reduction to machine-readable form (machine operation);*
- *Computer processing.*

Needs applicable to all sites are discussed first, and then special requirements, if any, for each of the five sites are discussed.

Quality control and confidentiality though distinct have a common requirement to maintain a controlled data base. Quality control assures the integrity of the data base by developing monitoring techniques and procedures that prevent the loss of all or part of any information. This assurance generally results from good control and monitoring procedures that permit the location on demand of any document wherever it may be in the processing cycle. Confidentiality needs this control to prevent the compromise of any data by limiting inadvertent or intentional access to the controlled data. Further, such controls permit the detection of any loss at the soonest possible time together with the information of who last held the document and was responsible for it.

The performance site should be specifically stated and in terms that delimit the boundaries for moving the data. Each site should have one individual responsible for confidentiality and inquiries about the site. The precision in stating the site location provides the basis for the remaining discussions. It is self-defeating to state the performance site as being the seventh floor, for example, when there exists no physical barrier to prevent access to the seventh floor from anywhere else in the building. To state the performance site as being the building does state more accurately the actual location but vastly enlarges the area to be protected. The Office of Toxic Substances must insure that it has access to the site for inspection in order to determine compliance with its requirements.

The Office of Toxic Substances must determine and promulgate the procedures for controlling access to its site and any remote site. Access control procedures will define the boundaries, material, documents, personnel, and accountability rules in force at each site. The boundaries (e.g., one room set aside as a restricted area) should be clearly marked and made known to all personnel. The material and documents subject to control should be clearly marked and identified. The extent of control reflects a cost decision which must be dealt with in an evolving system as an incremental process. Personnel must be trained and selected with needs of controlled access in mind. Again such training will evolve as the system evolves from a manual system to a computer data base. Accountability requires the designing of the entire system in a way that each document becomes the responsibility of one individual at a particular point in time; every document is so assigned and a record of the assignment is kept in some form. A library check-out system is an example of such an accountability system; the range of sophistication of library checkout systems indicates the range of choices available.

The performance site should provide the necessary storage facilities consistent with the media being used. The storage facilities should provide protection for the material both during the workday and after hours. Furthermore, the protection should extend to the physical facilities as well as to confidentiality of data. Loss through fire would impact the integrity of the data base as well as provide a mask for the compromising of information. Storage facilities often have inherent flaws which render them less than desirable. For example, basement or sub-surface storage is more vulnerable to flooding than above ground storage.

The mail room will require a special accounting function in order to process fees payable to EPA, once operations are underway.

Microfilming operations will require special storage capabilities for the film media; data entry and computer processing will also require several special storage capabilities.

f. Indexing

Although the annual reports from manufacturers, importers, and processors are amenable to structuring through good forms design and careful administrative control, the test data file will probably contain many forms and structures. Every attempt should be made to develop Test Protocols that necessarily lead to structured test data reports, but the manufacturer is not limited to any required input structure and may submit test data results in any conceivable form, even including reports or citations from the open literature. Though immediate requirements do not include abstracting and indexing, future plans should include the construction of an abstract for ease of dealing with the data base, the assignment of descriptors which serve to provide the necessary retrieval capabilities, and the posting of these descriptors to the various cataloging files. All of these tasks are highly intellectual tasks requiring specific skills and training in the areas of toxicology, chemistry, and information sciences. This assertion does not mean that the personnel required represent a small *or expensive subportion of the employment force. The experience of the New York Times Information Bank and the Institute for Scientific Information show that rather ordinary people can do the abstracting task quite adequately and, while special technical skills are needed to abstract technical subjects, the skill required is not extraordinary.*

B. Comparative Cost/Benefit Analysis

1. Cost Estimates

Based on the reported requirements and the possible systems offered as satisfying these requirements, the following cost estimates can be used to compare the options. At the same time it must be recognized that cost alone cannot be the determining factor, and that good planning requires a consideration of many additional criteria relevant to a task as important as the Office of Toxic Substances data base. The figures shown reflect estimates for the years 1975 through 1980. Additional comments are offered for the period beyond 1980.

a. Report Page Estimates

In order to compute the page count of the average report for use in estimating costs, the following estimates were made. Since the respondents would report 10,000 production items on 500 chemicals, the assumption is made that each respondent will report on one chemical substance in one report. Further, we assume that for each annual report there will be one page for the manufacturer identification and certification and one page of each chemical and its end-uses, by-products, and production figures.

For 1975 this means the 10,000 reports will have 20,000 pages. Similarly, the number of pages for each subsequent year's reports will be computed by doubling the number of production reports received each year. The total page volume is presented in Table 5 for each of the six years in the period 1975 through 1980.

b. Report Character Estimates

Estimates for the number of characters in an average report are based on the data elements listed in Exhibit 2. An average annual report is estimated to require a total of 4014 characters distributed among the following 5 items:

° Manufacturer	316 x 1 =	316
° Chemical Substance	265 x 1 =	265
° Production	138 x 1 =	138
° End-use	119 x 25 =	2975
° By-products	80 x 4 =	320
Total Characters		4014

An average premarket report is estimated to require a total of 780 characters distributed among the following 4 items:

° Manufacturer	316 x 1 =	316
° Chemical Substance	265 x 1 =	265
° End-use	119 x 1 =	119
° By-products	80 x 1 =	80
Total Characters		780

Annual Costs of Report Processes

	1975		1976		1977		1978		1979		1980	
	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High
1. Annual Reports	10,000		20,000		30,000		50,000		70,000		90,000	
2. Number of Character*	40,140		80,280		120,420		200,700		280,980		361,260	
3. Number of Pages	20,000		40,000		60,000		100,000		140,000		180,000	
4. Premarket Reports	500		500		500		500		500		500	
5. Number of Character*	390		390		390		390		390		390	
6. Number of Pages	1000		1000		1000		1000		1000		1000	
7. Total Character*	40,530		80,670		120,810		201,090		281,370		361,650	
8. Key punching Costs ($.30-$.50/1000) (130-Char/min)	$12,159 2.6 m/yr	$20,265	$24,201 5.2 m/yr	$40,335	$36,243 7.7 m/yr	$60,405	$60,327 12.9 m/yr	$100,545	$84,411 18.0 m/yr	$140,685	$108,495 23.2 m/yr	$180,825
9. Total Pages	21,000		41,000		61,000		101,000		141,000		181,000	
10. Microfilm Costs (.25 to .50/page)	$5,250	$10,500	$10,250	$20,500	$15,250	$30,500	$25,250	$50,500	$35,250	$70,500	$45,250	$90,500

* in 1,000's

Table 5

c. Other Estimates

- Keypunching
  $.30 to $.50 per 1000 keystrokes
  130 to 150 keystrokes per minute
- Microfilm
  $.25 to $.50 per page
- Microfilm Viewer with Automatic Retrieval
  $1,500 to $6,000
- Computer Processing Time
  $350 to $375 per hour
- Programming/Consulting Rates
  $15 to $30 per hour

All figures are rough approximations based on published material or previous experience.

2. Comparative Analysis

The following presentation compares the cost/benefits of the manual system with the computerized system. In areas where appropriate as a result of the comparison, the enhanced manual system is considered.

a. Indexing Point Capacity

Manual System

Pro's: The manual system, because of its inherent limitations, uses fewer people at lower levels to maintain the six indexing points proposed. Similarly, the start-up cost for initiating the various catalogs is negligible.

Con's: Because of the high risk of error in manual operations, particularly in high volume tedious work, the manual system may produce intolerable errors in the OTS data base through faulty catalog posting, filing, and the necessity for hand computations. The number of indexing points is inflexible and even the addition of one more indexing point to the existing data base is a significant task. The need to perform coordinated searches necessarily implies "pulling" two or more catalog cards to perform searches, and compounds the problem of maintaining data base integrity. Six reference points must be considered a restricted or limited search capability. Refiling pulled cards is as prone to error as the original filing. The manual system necessarily becomes an in-house capability because it manipulates the original documents and opportunities to sub-contract functions becomes more difficult to find.

Computerized System

Pro's: The computerized system's indexing points are almost unlimited and all fields in the file can become indexing points, at increased processing costs. The computerized system offers automatic catalog posting in the sense that catalogs become unnecessary or, at least, invisible to the users. Users work with copies of the file rather than with catalog cards and therefore remove nothing from the indexing catalogs which might interfere with the searches of other users. Though computers, too, can commit blunders, the mechanized system repeats proven functions accurately and is much less prone to a multiplicity of minor errors. The computerized system lends itself to sub-contracting the indexing and posting tasks since copies of data can be generated and combined easily.

Con's: The indexing points and the terms contained therein require careful coding since the computer is very literal and sees differences between terms like sodium chloride and salt.

b. *Input lag*

Manual System

Pro's: Since the bulk of the reports are annual reports the amount of time between receipt and posting of incoming reports is not critical. However, premarket reports with a response time of 90 days require a priority system to insure compliance.

Con's: In a manual system, overview and tickler file functions necessarily are manual, too. The onerous administrative functions must be borne by the clerical staff.

Computerized System

Pro's: The computerized system has the capability to provide any degree of currency with parallel increase in costs. In addition to this, once the data base is in machine-readable form a variety of administrative reports can be prepared to permit management to monitor the data base's currency.

Con's: All input must be reduced to machine-readable form.

c. *Response Time*

Manual System

Pro's: The system proposed provides an adequate response time for a very limited set of queries. In most cases the users may perform their own data base searches.

Con's: The preparation and performance of any complex query, can stretch into several days.

Computerized System

Pro's: The response time can be reduced to any future needs of OTS. The proposed system will provide 24 to 48 hour response time for *a virtually unlimited set of queries including computational and complex boolean search capabilities.*

Con's: None.

d. Data Extensibility

Manual System

Discussion: The manual system offers little help in restructuring the data files to take advantage of future goals. Any reordering may require manual sorts with that file unavailable for accessing during the sort.

Computerized System

Pro's: Once computerized, the data base can be extended, restructured, reorganized, and expanded in many different ways. This attribute should be particularly desirable if the data base and its usage are subject to increasing demands and evolving needs. New fields can be added trivially and references to data can reflect information about the data, (for example, "not collected prior to 1976", etc.).

Con's: The flexibility described above can become its own worst enemy and knowledgeable control must be exerted to prevent the destruction or confusion of information already in the data base.

e. Query Extensibility

Manual System

Pro's: Queries can always be processed by onerous serial searches of the data base.

Con's: The indexing points anticipate the possible queries which may be addressed to the data base but subsequent changes to them are difficult and require significant expenditure of clerical effort.

Computerized System

Pro's: As in the data extensibility discussion above, the computerized system gives great flexibility to query responses to anticipated questions but even novel searches requiring serial processing of the entire data base can still be accomplished in the 24 to 48 hour response time required by OTS.

Con's: Query construction requires specialized skills but technical users can be trained to perform fairly complex searches.

f. Reporting Facilities

Manual System

Discussion: The manual system offers no aids to the preparation of the reports resulting from searches on the data base.

Computerized System

Pro's: The computerized system offers a variety of reporting capabilities. At the same time the data base is searched, the computer can perform computational tasks, formatting, cross-footing of columns, pagination and high-lighting or summarization of the entire data base.

Con's: As in the preparation of computer queries, the structuring of computerized reports requires specialized skills in complex cases, and training of OTS personnel in the simple queries entered directly by users.

g. Availability of System Components

Manual System

Discussion: The manual system offers no problems in finding the equipment, personnel, or the development of procedures.

Computerized System

Pro's: The proposed computerized system offers all off-the-shelf components. There is no development time necessary for the basic software. The necessary hardware can be provided through utilization of the existing in-house computer installation, of computers at other EPA sites, or through contractor services. Aside from OTS's desire to keep the archival files stored within OTS's on-site storage facilities, the computerized system may be permanently sited at the contractor site if the proper precautions are provided.

Con's: Development lead time is required for the OTS application programs. These programs will be prepared in a high-level programming language and will undoubtedly be very straight forward and require less than two weeks for each such application or query. In fact, some simple queries could be prepared in a day. However, some problems may require significant lead time because of the problem complexity.

h. Personnel Constraints

Manual System

Discussion: The manual system because of its inherent limitations requires only clerical level help with skills in library or information sciences. Though experience in toxicological science will be useful to these people, undoubtedly an adequate level of training can be accomplished in-house.

Computerized System

Pro's: The computerized system replaces the need for much of the clerical labor required to perform the basic requirements of the

OTS. However, the staff of computer oriented personnel needed will be significantly higher paid than the clerical staff it replaces. Though more expensive, the skills necessary should not be difficult to acquire. Alternatively, with a computerized system the ability to sub-contract tasks becomes feasible. The trade-off of replacing clerks with programmers should always be thought of as necessarily unbalanced. The programmer is a toolmaker who builds tools for tasks that would be infeasible with the same dollar's worth of clerks. Furthermore, the task can be repeated over and over again with the same accuracy; it can be done at speeds unthinkable with clerks; and it can be done at anytime, day or night.

Con's: The requirement to maintain the archival file of reports necessitates the parallel structure of a manual file even if the computerized approach is taken. Further the archival data base must also reflect the administrative actions taken by the OTS, such as, including notification of non-compliance, for example.

i. Security

Manual System

Pro's: The simplicity of the manual system offers the easiest path to a totally secure system. As the Section on confidentiality states, the entire file is placed in a secure storage place and submitted to security procedures similar to those applied to any valuable item.

Con's: Manual systems are prone to errors, undetected losses, misfiles, and other human-related frailties which may mask or even create the opportunity for the compromise of confidential data.

Computerized System

Pro's: The computerized system can provide many supervisory or monitoring capabilities that facilitate the detection of unauthorized access, document losses, and improper data insertions. These functions can be performed in conjunction with the normal operation of the system or quickly and accurately performed for ad hoc purposes, a capability impossible with manual systems.

Con's: In contrast to the simplicity of the manual system, the complexity of computer systems (especially, time-shared systems using communications lines) makes any assurance of complete computer security very difficult.

3. Comparative Costs

The comparative costs for the systems proposed are presented in the following tables. The costs of the enhanced manual system are discussed under the cost/benefit analysis of each function. Costs associated with overhead have not been included in any of the estimates since they remain fairly constant across the various phases. These expenses include such items as forms printing, office equipment, floor space rental, supplies and photocopying.

Table 6 presents the personnel costs for the manual system over the six year period 1975 to 1980. Personnel costs represent the total cost associated with the manual system.

Table 7 presents the personnel costs associated with the computerized system. The personnel costs are entered into Table 8 to provide the total costs for the computerized system. It should be noted that the personnel described may be within the staff of the service organization providing the functions.

Phase I Personnel Roster and Costs 1/

	1975		1976		1977		1978		1979		1980	
Facility Manager (GS-12)	1	$ 18,000	1	$ 18,000	1	$ 18,000	1	$ 18,000	1	$ 18,000	1	$ 18,000
Technicial Information Specialist (GS-7)	1.5	15,000	1.5	15,000	1.5	15,000	1.5	15,000	1.5	15,000	1.5	15,000
Data Technicians (GS-5)	.5	4,500	1	9,000	1.25	11,250	2	18,000	3	27,000	4	36,000
Library Technicians (GS-4)	1.75	14,000	3	24,000	4.75	38,000	7.5	60,000	10	80,000	13	104,000
Record Clerks (GS-3)	.5	3,500	1	7,000	1.25	8,750	2	14,000	3	21,000	3.5	24,500
File Clerks (GS-3)	.5	3,500	.5	3,500	.5	3,500	1	7,000	1.5	10,500	1.5	10,500
TOTALS	5.75	$58,500	8.0	$76,500	10.25	$94,500	15	$132,000	20	$171,500	24.5	$208,000

1/ Personnel figures are stated in person-years.

Table 6

Phase III Personnel Roster and Costs 1/

	1975		1976		1977		1978		1979		1980		
Facility Manager (GS-13)	1	$ 21,000	1	$ 21,000	1	$ 21,000	1	$ 21,000	1	$21,000	1	$ 21,000	
Systems Analyst (GS-12)	.5	9,000	.5	9,000	1	18,000	1	18,000	1.5	27,000	1.5	27,000	
Programmer (GS-12)	.5	9,000	.5	9,000	1	18,000	1	18,000	1.5	27,000	1.5	27,000	
Technical Information Specialist (GS-7)	1	10,000	1	10,000	1	10,000	1.5	10,500	1.5	10,500	1.5	10,500	
Data Technicians (GS-5)	.5	4,500	1	9,000	1.25	11,250	2	18,000	3	27,000	4	36,000	
Library Technicians (GS-4)	.5	4,000	1.25	10,000	2	8,000	3	12,000	4.25	34,000	5.5	44,000	
Keypunch Operators (GS-4)	2.6	20,265	5.2	40,335	7.7	60,405	12.9	100,545	18	140,685	23.2	180,825	2/
Record Clerks (GS-3)	.5	3,500	.75	5,250	1	7,000	2	14,000	2.75	19,250	3.5	24,500	
File Clerks (GS-3)	.5	3,500	.5	3,500	.5	3,500	1	7,000	1.5	10,500	1.5	10,500	
Microfilm Clerks	.5	3,500	1.5	10,500	2.5	17,500	4	28,000	5.5	38,500	7.5	52,500	
TOTALS	8.1	$88,265	13.2	$127,585	18.95	$174,655	29.4	$247,045	40.5	$355,435	50.7	$433,825	

1/ Personnel figures are stated in person-years.

2/ From Table 5, Annual Costs of Report Processes

Table 7

a. Computerized System

Data base maintenance costs for the computerized system are based on a weekly pass of the entire data base on magnetic tape as a batch process. The parameters used are: a magnetic tape reading speed of 45,000 characters per second; and computer processing costs of $375 per hour. Reducing the periodic processing rate to monthly would reduce the cost to one quarter of the values shown.

Information Retrieval costs necessarily reflect the number and complexity of the questions applied to the data base. Many of the recurring queries can be processed during the file maintenance processes. However, an arbitrary one hour of processing time per month is included in the cost estimates for special queries each month. In a batch system, query processing time may increase with the growth of the data base, but this factor was ignored.

The costs for archival file maintenance are taken from the manual system costs. The costs are based on a judgement that the filing and accession, control, and maintenance of the archival file can be accomplished by three record clerks.

Computerized Systen Costs

	1975	1976	1977	1978	1979	1980
Report volume						
1. Annual	10,500	20,500	30,500	50,500	70,500	90,500
2. Daily	42	82	122	202	282	362
File Storage						
3. Pages	21,000	62,000	123,000	224,000	365,000	546,000
4. Characters(1,000's)	40,530	121,200	242,010	443,100	724,470	1,086,120
Personnel Costs	$88,265	$127,595	$174,655	$247,045	$355,435	$433,825
Computer Processing						
Data Base Maintenance	$ 4,878	$14,589	$29,130	$53,336	$ 87,204	$130,736
Info Retrieval	$ 4,500	$ 4,500	$ 4,500	$ 4,500	$ 4,500	$ 4,500
Microfilming Cost	$ 7,875	$15,375	$22,875	$37,875	$ 52,875	$ 67,875
TOTAL						

Table 8

b. Enhanced Manual System

In comparing the rather basic manual system against a computerized system dependent on the full reduction of annual and premarket reports to machine-readable form, the recurring theme is the low cost of the limited manual system against the flexibility of the expensive computer system. The enhanced manual system described various options which permitted the use of the computer to extend or expand the manual capabilities without taking the full step to a computerized system. The following discussion gives a cost/benefit analysis of each enhancement described the Section III(B).

1. Indexing Aids

The benefits appreciated through this approach over the manual system include the savings inherent in the avoidance of several typing and proofreading operations on each document indexed. Final hardcopy can be produced on a variety of devices with an accuracy and cleanness not possible with typescript. Subsequent corrections can be applied to the basic document without re-typing the entire document and yet producing a clean hardcopy. A further benefit is the development of a machine-readable data file which can later be used in the OTS Data Base or, with suitable data conversion, in existing data bases such as TOXLINE. Costs range from approximately $4500 for a Datapoint 2200 minicomputer to several dollars of computer processing time from a terminal that can access System 2000. This approach could extend by several fold the number of input reports processed by one clerk. This approach would involve a negligible cost increment.

2. Catalog Posting and Publication

Cost considerations could never justify this approach, particularly in a slowly growing data base. Emphasis must be placed on the capability of having a very large number of indexing points and the abailability of other computer oriented tools, such as Key-Word-In-Context listings. Computer processing costs will actually increase exponentially as the total number of documents in the data base increases linearly. However, the tasks of assigning new indexing points, regrouping old indexing points, assigning synonyms, and reordering the minor sort-keys can be accomplished with very little additional cost at the next periodic catalog preparation.

The factors affecting the cost of this approach include the indexing of incoming reports onto a machine-readable form. This value is a direct function of the number of indexing points selected and the consequent number of characters required to be keystroked. The skills required exceed that of an ordinary keypunch operator and will undoubtedly require in-house training. The processing costs necessary to sort and selectively print the various catalog listings are another cost factor. The keystroking of the catalog indices is estimated at one tenth the cost of keystroking the entire report. Personnel costs are estimated to exceed other manual personnel costs by 30%. Processing time is estimated to be approximately one hour per month.

3. Data Descriptor Vocabulary

A major benefit is the control exerted on the ambiguities of the English language and the development of an organized way to address the problems of synonyms, classification of concepts, and neologisms. The indexers and future abstractors must continuously battle these problems and the periodic publication of the data descriptor vocabulary in dictionary form will assist them.

The cost of a controlled vocabulary is estimated at $4000 to $5000 a year in computer processing time. The benefit is too intangible to permit comparing a controlled vocabulary to an uncontrolled one.

4. Data Element Dictionary

A Data Element Dictionary provides a means to control the data files and a means of communication to other potential users of the data base. It eases eventual computerization by formalizing data fields.

5. Bibliography Preparation

The cost of maintaining an updated list of test data documents at the estimated input volume is approximately $4000. The cost of producing this program from existing generalized programming aids is $3000. The utility of such ad hoc special lists rests in the ability to satisfy unanticipated needs.

6. Automated Mailing List and Tickler File

The cost of programming using generalized programming tools is estimated to be $4000. The cost of each run will be a direct function of the number of respondents in the file. At the rate of 10,000 to 90,000 annual reports estimated for the six year period, the annual cost each year would be less than $10,000.

7. Microfilm Working Files

Microfilming has been justified in several installations in terms of the storage space saved, the insurance provided against catastrophe, savings over paper reproduction costs, and reduced office space requirements. Such an analysis awaits more precise data of the actual form of many of the documents OTS will be receiving. However, the primary benefit to be emphasized here is its potential for union with future computer systems. Microfilming costs .25 to .50 per page for the master film. First year costs are estimated to be $7000 to $10,000. The sixth year costs which may be considered indicative of the annual rate are estimated as approximately $45,000 to $90,000.

8. Microfilm Retrieval Systems

Such a system would be primarily intended for the test data file which is unlikely to ever be fully inputted into a computer retrieval system. Computerized bibliographic search systems, if implemented, would augment the benefits achieved from such a system.

C. Guidance for RFP

1. Recommendation for Standard Proposal Outline

The Request for Proposals under consideration by the Office of Toxic Substances will seek to satisfy the data management needs under the proposed Toxic Substances Control Act.

By seeking contractor services to provide the administrative and clerical functions associated with the extremely important toxic substances data base, OTS will free its staff for the oversight and regulatory responsibilities under the proposed Act. Since the basic responsibilities must remain with OTS, every effort must be made to insure that the ultimate vendor can, indeed, meet his delegated responsibilities.

Because of the unique nature of the new EPA responsibility, the uncertainty of tasking parameters, the possible changes in the proposed legislation before passage, and the almost immediate production requirements of the legislation, the Office of Toxic Substances must seek a contractor who can satisfy a variety of possible approaches to this task. At the outset, the requirements of the task may not justify full time assignments of personnel. However, the continued growth of the data base is assured and the contractor must have the ability to meet the growing demands. Further, the contractor must have the experience needed to contribute to the development of enhancements and extensions to the evolving system.

The Request for Proposals must require all bidders to respond in conformance with the standard proposal outline. Adherence to this requirement will facilitate proposal evaluations, insure complete responses, and satisfy the need for assuring each bidder of impartiality.

The content of the standard proposal outline is intended to provide OTS with the necessary information to evaluate the ability of the bidder to perform adequately the tasks proposed. The weights assigned to the various sections indicate NBS recommendations for the relative importance of each of the sections.

Standard Proposal Outline

I. Statement of the Problem 5%

II. Overview of Suggested Processing Flow 10%

- A. Initial Manual System
- B. Potential Manual System Enhancements
- C. Computerized Information Processing System

III. Proposed Manual System 20%

- A. Forms Design
    1. Source of Services
       (in-house, sub-contracted, etc.)
    2. Performance Site
       (actual site where function is performed)
    3. Facilities Available
       (description of equipment, procedures, and special skills required for this function)
    4. Personnel Assigned
       (statement of individuals assigned and percent of time required)
    5. Confidential Provisions
       (statement of confidentiality provisions specific to this function)
- B. Mail Receipt and Accession Control 1/
- C. Report Preparation, Compilation, Binding, and Dissemination 1/

IV. Potential Manual System Enhancements 10%

- A. Microfilm 1/
- B. Reduction to Machine-readable Form 1/

V. Computerized Information Processing System 20%

1/ The subsections in III(A) are repeated here.

A.	*Computer Processing in Support of Manual System Enhancements* 1/	
B.	*Generalized Data Management System Package* 1/	
VI.	*Proposed Confidentiality and Security Procedures*	15%
A.	*Contractor statement on the ownership of all data developed under this contract: compiled lists; catalogs; reports; computer programs; and all storage media.*	
B.	*Detailed Plan for Meeting Confidentiality Requirements*	
VII.	*Personnel*	5%
A.	*Organization Plan (describe relationships, reporting channels, and relationship with the Office of Toxic Substances)*	
B.	*Resumes of all Non-clerical Personnel (include relationship with contractor [full-time employee, consultant, etc.], length of service with contractor, pertinent experience, and position for which offered)*	
VIII.	*Corporate Experience with Similar Tasks*	8%
IX.	*Government Supplies or Services Required by the Contractor*	5%
X.	*Equipment Necessary to support Proposal*	2%

1/ Subject to final determination by OTS.

2/ This entire form is repeated for the first year, second year, and third year costs as major headings A, B, and C respectively.

XI. Cost Figures

A. Annual Costs 2/	A. First Year	B. Second Year	C. Third Year
1. Parameters used for costs			
Annual Reports	1/		
Premarket Reports	1/		
Test Data Reports	1/		
Activity against data base			
Premarket queries			
Summary Reports			

2. Anticipated Costs Under the Proposed Manual System

Forms Design $__________

Mail Receipt and Accession Control $__________

Technical Analysis $__________

Report Preparation, Compilation, etc. $__________

3. Enhancement Costs

Microfilming

Master film cost per page ________

Annual Costs $__________

Working or retrieval copy/page ________

Annual Costs $__________

Reduction to Machine-Readable Form

Cost per 1000 characters of data________

Annual Costs $__________

4. Computer Processing Costs

Enhancement Support Costs

Computer Processing Costs

Per hour ______

Annual Costs $______

5. Data Base Processing Costs

Computer processing Costs

Per hour ______

Annual Costs $______

On-line Storage Costs

Appropriate unit cost ______

Annual Costs $______

Software Development Costs (annual) $______

List of Products Developed

6. Equipment Costs

List all equipment required under the proposal $______

V. RECOMMENDATIONS

A. Overview

Current estimates of reports that will be received as a result of the regulations issued by the Office of Toxic Substances reveal that by 1980, OTS will be receiving an annual influx of 90,500 reports (or 362 a day) and will be dealing with a file of over a half million pages of data. This, combined with the planning constraint of limiting OTS's participation in this work to four people over this six year time frame, strongly suggests that computerized assistance will be mandatory by 1980. Furthermore, there is no reason to assume that the 1980 input report volume will not continue to grow in following years. However, the initial volume of 10,500 reports (or 40 a day) does not justify the use of automation.

Therefore, it is recommended that the manual system be installed with a life time expectancy of four years (1975-1978) and that this time be used to gain experience with real data, collect statistics, reaffirm the estimates, develop assessments of the available encoding schemata, and "shakedown" the reporting form design. However, the planning for the eventual automation should begin at once with services sought from EPA's Management Information and Data Systems Division on appropriate standardization and development plans for the enhancements to the manual system. It is also recommended that OTS use the on-board computer system and the System 2000 Data Management System for the enhancements. Both of these systems are currently in use in EPA's Management Information and Data Systems Division.

Though experimental or pilot programs are encouraged, production keystroking and microfilming should be delayed until the third year when the aggregate of the three years will offer sufficient volume for subcontracting (206,000 pages to microfilm, 404 million characters to keystroke). Microfilming cannot be justified if actual experience shows few retrievals of the annual reports. Microfilming of the test data file does not seem justified on the basis of current volume and requirements.

B. Recommended Program

1. First Year (1975)

With assistance (possibly of the Bureau of the Census), design a reporting form which supports and explains the informational goals of the published regulations. Initiate the manual system by hiring the identified skills. Begin a monitoring program to develop statistics of actual experience with respondents and OTS data users. Collect and analyze end-use responses for refinement of reporting requirements and possible development of other end use codes. Begin collection of data descriptors to provide the basis for a controlled vocabulary. Investigate possible keystroking and microfilming services.

2. Second Year (1976)

Continue use of manual system, statistics collection, and monitoring of actual experience. Resolve end-use coding question. Develop experimental programs for microfilming formats *if* test data file growth

PROGRAM PHASING

1975	1976	1977	1978	1979 - 1980
Design Report form	Initiate Data Element Dictionary			
Initiate Manual System			Manual System in Parallel with Computerized System	
Establish Monitoring Program			Review User Requirements	
Investigate End use codes	Complete End use Study			
Begin manual Collection of Data Descriptors				
Investigate Keystroking and Microfilming Services	Perform Experimental Microfilming	Initiate Production Keystroking		
		Review Microfilm		
	Initiate Automated Catalog Posting			Develop Library of Recurring Queries
		Initiate Automated System	Automated System in full operation	

Figure 6. Program Phasing

justifies it. Begin development of automated catalog posting and publishing in parallel with manual system posting. Develop automated indexing aids when catalog posting process is operating fairly reliably. Expand the number of indexing points. Begin development of Data Element Dictionary.

3. Third Year (1977)

Continue use of manual system, statistics collection, and monitoring of actual experience. Review microfilm usage for test data and for reports on basis of retrieval activity. Begin production keypunching of annual and premarket reports in parallel with manual system enhancement processes. Begin development of computerized system in parallel with manual system.

4. Fourth Year (1978)

Continue use of manual system in parallel with computerized system but set date for switch to full computerized system for beginning of fifth year. Begin intensive review of system satisfaction of user needs.

5. Fifth and Sixth Years (1979-1980)

Continue use of computerized system. Develop library of recurring queries and review completeness of data base. Review data base growth and activity.

C. Specific Topics

1. Chemical Substance Searching

By including the CAS Registry number in the manufacturer reports, the OTS will have the ability to search for one or more unique chemical substances in either the manual or computerized system. Chemical substructure searching will not be possible with any of the systems described. However, it is recommended that this capability be obtained through the use of existing EPA chemical searching capabilities as an auxiliary, independent operation until the utility of duplicating the chemical structure data within the OTS data base can be established.

2. Confidentiality.

Under the manual system the confidential data will be completely located within the OTS and protected as any valuable item should be. However, the series of recommendations made in the Confidentiality Section represents a significant effort which must be addressed. Under the computerized system, OTS must seek a totally dedicated system and must bear the subsequent expense.

3. Reporting Forms

OTS should seek competent assistance in the development of a reporting form which satisfies informational needs, manual and computerized processing considerations, and the task of reducing the reports to machine-readable form.

4. Interface with Existing Data Bases

OTS should participate in and encourage efforts by EPA Management Information and Data Systems to develop standards for data base management.

Appendix A. *Existing Bibliographic Data Bases*

The following list of bibliographic data bases which are pertinent to the needs of the Office of Toxic Substances was compiled from NBS Technical Note 814, "A mechanized Information Services Catalog." A description of the entries can be found in the parent document. 1

1. ABSTRACTS ON HEALTH EFFECTS OF ENVIRONMENTAL POLLUTANTS
2. INDEX TO API ABSTRACTS OF REFINING LITERATURE
3. INDEX TO API ABSTRACTS OF REFINING PATENTS
4. BIOLOGICAL ABSTRACTS REVIEWS
5. CA INTEGRATED SUBJECT FILE
6. CHEMICAL ABSTRACTS SERVICE SOURCE INDEX
7. CHEMICAL ABSTRACTS CONDENSATES
8. CHEMICAL-BIOLOGICAL ACTIVITIES
9. CHEMICAL MARKET ABSTRACTS TAPE
10. CHEMICAL TITLES
11. COMPREHENSIVE DATA BASE OF PATENTS
12. COMPUTER BASED NUCLEAR MAGNETIC RESONANCE LITERATURE RETRIEVAL SYSTEM
13. COMPUTERIZED INFORMATION RETRIEVAL SYSTEM OF THE GAS CHROMATOGRAPHY LITERATURE
14. CPI MAGNETIC TAPE FILE
15. CURRENT PROGRAMS
16. EXCERPTA MEDICA
17. FOOD SCIENCE & TECHNOLOGY ABSTRACTS
18. GEOLOGICAL REFERENCE FILE
19. GOVERNMENT REPORTS ANNOUNCEMENTS
20. INDEX CHEMICUS REGISTRY SYSTEM
21. INFORMATION SYSTEM
22. INTERNATIONAL TREE DISEASE REGISTER
23. IOWA DRUG INFORMATION SERVICE
24. ISI CITATION MAGNETIC TAPES
25. ISI SOURCE INDEX MAGNETIC TAPES
26. MASS SPECTROMETRY BULLETIN
27. METALS ABSTRACTS INDEX
28. PANDEX CURRENT INDEX TO SCIENTIFIC AND TECHNICAL LITERATURE
29. ABSTRACT BULLETIN OF THE INSTITUTE OF PAPER CHEMISTRY
30. P.A.S.C.A.L.
31. PATENT CONCORDANCE IN COMPUTER - READABLE FORM
32. PATENT OFFICE MECHANIZED SEARCH SYSTEMS
33. PESTDOC
34. PETROLEUM ABSTRACTS MASTER RECORD TAPES
35. POLYMER SCIENCE AND TECHNOLOGY
36. RINGDOC
37. S.A.B.I.R.
38. TEXTILE TECHNOLOGY DIGEST KEYTERM INDEX

39. *TOXICOLOGY INFORMATION OCNVERSATION ON-LINE NETWORK*
40. *UNION CATALOG OF MEDICAL PERIODICALS*
41. *VETDOC*